ABSTRACT

YAN, HONGQIANG. Essays on High-Dimensional Threshold Models. (Under the direction of Mehmet Caner).

This dissertation focuses on high-dimensional threshold models, particularly cases where the number of regressors may exceed the sample size. In the first essay, I explore a regression model with a possible change point due to a covariate threshold and propose a debiased (or desparsified) LASSO estimator designed for accurately estimating high-dimensional threshold regression models. In the second essay, I apply the methodology developed in the first essay to investigate fully flexible, nonlinear threshold models of price linkages in spatially distinct world maize markets.

The first essay addresses statistical inference for high-dimensional threshold regression parameters. I establish oracle inequalities for the scaled LASSO estimator proposed by Lee, Seo, and Shin, assuming only non-subgaussian error terms and covariates. Subsequently, I debias (or desparsify) the scaled LASSO estimator and derive the asymptotic distribution of tests involving an increasing number of slope parameters, following the approach outlined in van de Geer et al. (2014). Utilizing these results, I construct asymptotically valid confidence intervals for the components of the threshold regression slope coefficients. To complement the asymptotic theory in this paper, I conduct simulation studies to demonstrate the performance of my method in finite samples.

Threshold models have also played a significant role in studying price transmission in agricultural economics. The second essay investigates the extent of market integration and exchange rate pass-through, as well as market factors that may be associated with deviations from perfect market integration and pass-through. To address the shortcomings of existing models of spatial market integration, I employ procedures outlined in the first chapter. My results support the integration of world maize markets, especially when accounting for the existence of thresholds. I identify significant relationships among several variables representing domestic and world economic conditions.

Essays on High-Dimensional Threshold Models

by
Hongqiang Yan

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Economics

Raleigh, North Carolina
2024

APPROVED BY:

Barry Goodwin	Zheng Li
Ilze Kalnina	Mehmet Caner Chair of Advisory Committee

BIOGRAPHY

The author was born in the Xinjiang Uygur Autonomous Region, People's Republic of China. Before pursuing the PhD Degree at North Carolina State University, he earned a bachelor's degree in finance from Beijing Jiaotong University.

ACKNOWLEDGEMENTS

I take this opportunity to express my deepest gratitude to all those who have supported and guided me throughout my doctoral journey and the completion of this dissertation.

First and foremost, I extend my heartfelt appreciation to my advisor, Dr. Mehmet Caner, for his invaluable guidance, mentorship, and unwavering support. Dr. Caner's expertise and encouragement have played a pivotal role in shaping the major research question and opening new doors of exploration. Additionally, I am grateful for the guidance from other committee members, particularly Dr. Barry Goodwin. Dr. Goodwin consistently encouraged me to translate complex results into meaningful economic insights and assisted in conducting applied economic research. Their expertise, dedication, and constant encouragement have been instrumental in the successful completion of this dissertation. I extend my gratitude to committee members Dr. Zheng Li and Dr. Ilze Kalnina for their time and invaluable insights, which greatly enriched my ideas and methodologies through their thorough review of my work. Special appreciation goes to Dr. Xiaoyong Zheng and Dr. Roderick Rejesus for their invaluable advice and collaboration throughout my PhD journey.

I am deeply grateful to my family for their love, support, and understanding throughout this challenging process.

I acknowledge the financial support from the Goodnight Doctoral Fellowship for my last two years, provided by NC State University in partnership with Dr. Jim and Mrs. Ann Goodnight and the Goodnight Educational Foundation. I also appreciate the Jenkins Fellowship provided by the Poole College of Management at NC State University, enabling me to focus on my doctoral studies and dissertation during the final year. Finally, my heartfelt thanks to the entire faculty and staff of the Department of Agriculture and Resource Economics in the College of Agriculture and Life Sciences and the Department of Economics in the Poole College of Management, whose unwavering support and encouragement have been invaluable throughout my academic journey.

Thank you for being an essential part of my academic endeavor.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER

1

UNIFORM INFERENCE IN HIGH-DIMENSIONAL THRESHOLD REGRESSION MODELS

1.1 Introduction

Threshold models are a popular way to characterize nonlinearities in economic relationships. Hansen (1996) and Hansen (2000) show how the least squares estimation of threshold models is possible and feasible in fixed-dimensional settings, where the number of observations is much larger than the number of variables. These two papers develop a non-standard asymptotic theory of inference which allows for the construction of confidence intervals for the regression estimates, as well as testing of hypotheses for the presence of a threshold. Later, Caner and Hansen (2004) developed instrumental variable estimation techniques that allow for the covariates to be endogenous.

Let $\{(Y_i, X_i, Q_i) : i = 1, \ldots, n\}$ be a sample of independent observations such that

(1.1.1) $$Y_i = X_i'\beta_0 + X_i'\delta_0 \mathbf{1}\{Q_i < \tau_0\} + U_i, \quad i = 1, \ldots, n,$$

where for each i, X_i is a $p \times 1$ vector, Q_i is a scalar, U_i is error terms, and $\mathbf{1}\{\cdot\}$ denotes the indicator function. The scalar variable Q_i is the threshold variable determining regime switching and τ_0 is the unknown threshold parameter.

Threshold models have been well studied and applied in econometrics. In empirical studies, threshold models have been used to investigate the non-linearity in the threshold effect of government debt on economic output(e.g. Chudik et al. (2017), Afonso and Jalles (2013), Grennes et al. (2010)). Recently, there has been a growing interest in panel threshold models. Seo and Shin (2016) propose a two-step GMM estimator for the dynamic panel threshold model, which also allows for the endogeneity of either the covariates or the threshold variables. Miao et al. (2020a) study estimation and inference in a panel threshold model in the presence of interactive fixed effects. Miao et al. (2020b) consider latent group structures in a panel threshold regression model, which allows for the slope coefficients and threshold parameters to vary across individual units.

Interest in high-dimensional data has motivated much recent research on LASSO for threshold regression. Lee et al. (2016) establish sparsity oracle inequalities for the prediction norm and estimation error of the scaled LASSO applied to (1.1.1) in the case of fixed regressors and Gaussian error terms for both the no threshold effect case and the threshold effect case. In their simulation section, they also extended their results to random regressors with Gaussian errors. Callot et al. (2017) develop sup-norm oracle inequalities for the estimation error of the LASSO of Lee et al. (2016). Then they propose a thresholded scaled LASSO estimator based on the sup-norm bound to provide threshold selection consistency or even model selection consistency.

Rapid technological advancements in data collection and processing have led to the analysis of high-dimensional datasets, where the number of variables far exceeds the sample size. In such high-dimensional settings, classical inferential procedures, such as maximum likelihood, are no longer valid. Consequently, there is a pressing need to develop new principles, theories, and methods for parameter estimation, hypothesis testing, and confidence intervals (CIs). Our approach is an adaptation of the debiasing a LASSO estimator introduced in van de Geer et al. (2014). Specifically, in van de Geer et al. (2014), a debiased LASSO estimator is proposed, and asymptotically valid confidence bands for the estimated parameters are constructed. Similar advancements were made in the papers by Zhang and Zhang (2014) and Javanmard and Montanari (2014). The idea is to remove the bias introduced by shrinkage by debiasing the estimator with a constructed approximate inverse of a singular sample covariance matrix for estimating high-dimensional regression models. Two approaches are widely used to construct the approximate inverse matrix: the nodewise regression introduced by Meinshausen and Bühlmann (2006) and the CLIME estimator of Cai et al. (2011).

Much of the present work is devoted to solving the inference problem in a high-dimensional linear model by debiasing LASSO-type estimators to construct asymptotically valid confidence bands for the parameters of interest. Caner and Kock (2018) propose the conservative LASSO estimator allowing for non-identically distributed or non-sub-Gaussian error terms and develop the asymptotic distribution of tests involving an increasing number of parameters. Gold et al. (2020) propose a debiased LASSO based on a two-stage least squares estimator with sub-Gaussian data and homoskedastic errors for a high-dimensional instrumental variables regression. They allow both the number of instruments and the number of regressors to be greater than the sample size. Another relevant paper is by Caner and Kock (2019), which develops a debiased GMM estimator for estimating a high-dimensional instrumental variables regression that has many more endogenous regressors than observations. In their simulations, they compare it to the estimator in Gold et al. (2020). Belloni et al. (2019) provide a new way of handling linear and nonlinear instrumental variables regression as well as relaxing the sparsity assumption.

The present work, as introduced by Belloni et al. (2014) and Semenova et al. (2021), proposes estimation methods for desparsification in treatment effects. In the context of generalized linear models, relevant articles include those by Belloni et al. (2016) and Caner (2023). Additionally, it's worth noting that high-dimensional time series models are considered in papers by Adamek et al. (2023). Moreover, high-dimensional panel data models are addressed in works by Kock (2016), Kock and Tang (2019), and Chiang et al. (2023).

Overall, I contribute to the literature in two ways. Our primary contribution is to develop a debiased LASSO estimator for the threshold regression in the high-dimensional regime: $p >> n$— that is, if p, the number of variables is much larger than n, the number of observations. The estimator in Lee et al. (2016) may be debiased in the sense of van de Geer et al. (2014) in order to construct asymptotically uniform confidence intervals for the parameters of interest and hypothesis tests under a sparse setting. Despite the considerable progress that has been made for inference in linear high-dimensional regression, only a few papers provide theoretical insights into more complex models, such as nonlinear models. Another contribution involves extending oracle inequalities for the LASSO estimator for high-dimensional threshold regression to non-subgaussian error terms and regressors, using the maximal inequalities by Chernozhukov et al. (2017). Strengthening our assumption of sub-Gaussianity could deliver even stronger results.

The rest of the paper is organized as follows. Section 1.2 recalls the LASSO estimator of Lee et al. (2016). In Section 1.3 I develop oracle inequalities for the LASSO estimator of regression slopes as well as the threshold estimator only assuming non-sub-Gaussian error terms and regressors. In section 1.4 I propose a debiased LASSO estimator for the high-dimensional threshold regression model and derive the asymptotic distribution of hypothesis tests for slope parameters based on an adaptation of the work in van de Geer et al. (2014). In section 1.5 I investigate the finite sample properties of the debiased LASSO for threshold models and compare it to the debiased LASSO estimator for linear models of van de Geer et al. (2014). All proofs are deferred to the Appendix.

1.2 The Model

Notation

For $\{(Y_i, X_i, Q_i) : i = 1, \dots, n\}$ following (1.1.1), let bold font $\boldsymbol{X}_i(\tau)$ denote the $(2p \times 1)$ vector such that $\boldsymbol{X}_i(\tau) = (X_i', X_i' \mathbf{1}\{Q_i < \tau\})'$ and let $\boldsymbol{X}(\tau)$ denote the $(n \times 2p)$ matrix whose $i-th$ row is $\boldsymbol{X}_i(\tau)'$. Let X and $X(\tau)$ denote the first and last p columns of $\boldsymbol{X}(\tau)$, respectively.

For any $L \times 1$ real vector a, let $\|a\|_q$ denote the ℓ_q norm of a. Particularly, if $a = (a_1, \dots, a_n)'$, n-dimensional vector, the prediction norm is defined as $\|a\|_n := \sqrt{\frac{1}{n} \sum_{i=1}^n a_i^2}$.

Also, let $J(a) := \{j \in \{1, \dots, L\} : a_j \neq 0\}$ and let $|J(a)|$ denote the cardinality of $J(a)$. Let $\mathcal{M}(a)$ denote the number of nonzero elements of a, i.e. $\mathcal{M}(a) = |J(a)|$. Then I let $b_J \in \mathbb{R}^L$ denote the vector has the same coordinates as a on J and zero coordinates on the complement J^C. Let the superscript $^{(j)}$ denote the $j-th$ element of a vector or the $j-th$ column of a matrix depending on the context.

For any $m \times n$ matrix A, I define $\|A\|_\infty := \max_{1 \le i \le m, 1 \le j \le n} |A_{ij}|$. $\|A\|_{l_\infty} := \max_{1 \le i \le m} \sum_{j=1}^n |A_{ij}|$ denotes the induced l_∞-norm of A. Similarly, $\|A\|_{l_1} := \max_{1 \le j \le n} \sum_{i=1}^m |A_{ij}|$ denotes the induced l_1-norm of A.

Finally, define $f_{(\alpha,\tau)}(x, q) := x'\beta + x'\delta \mathbf{1}\{q < \tau\}$, $f_0(x, q) := x'\beta_0 + x'\delta_0 \mathbf{1}\{q < \tau_0\}$, and $\widehat{f}(x, q) := x'\widehat{\beta} + x'\widehat{\delta} \mathbf{1}\{q < \widehat{\tau}\}$. Then, I define the prediction norm as

$$\left\|\widehat{f} - f_0\right\|_n := \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\widehat{f}(X_i, Q_i) - f_0(X_i, Q_i)\right)^2}.$$

Throughout the paper, I use the superscript zero to signify the true parameter value.

1.2.1 LASSO Estimation

We consider the model in (1.1.1). It can be written as

$$Y_i = \begin{cases} X_i'\beta_0 + U_i, & \text{if } Q_i \ge \tau_0, \\ X_i'(\beta_0 + \delta_0) + U_i, & \text{if } Q_i < \tau_0. \end{cases} \tag{1.2.1}$$

Q_i in the above model is used to split the sample into two groups. When $Q_i < \tau_0$, the regression function becomes $X_i'(\beta_0 + \delta_0) + U_i$; if $Q_i \geq \tau_0$, the regression function reduces to $X_i'\beta_0 + U_i$. As δ_0 is the change of regression coefficients between two regimes, the model in (1.1.1) captures a regime switch based on an observable scalar variable Q_i with a scalar unknown parameter τ_0. The case of $\delta_0 = 0$ corresponds to the linear model. If $\widehat{\delta} = 0$, then this case amounts to selecting the linear model.

Recall the model in Lee et al. (2016). Further assumptions in the model are detailed in Section 1.3. Let $\alpha_0 = (\beta_0', \delta_0')'$. Then, using notation defined above, I can rewrite (1.1.1) as

$$Y_i = \boldsymbol{X}_i(\tau_0)'\alpha_0 + U_i, \quad i = 1, \ldots, n. \tag{1.2.2}$$

α_0 is the $2p \times 1$ population vector of coefficients, which I shall assume to be sparse. However, the location of the non-zero coefficients is unknown and potentially $2p$ could be much greater than n. We assume that the explanatory variables are exogenous and precise assumptions will be made in Assumption 1 below. Let $J_0 = J(\alpha_0)$, denote the set of non-zero coefficients and $s_0 = |J_0|$, the cardinality. In this paper, I study the high-dimensional case where p is much greater than n.

Let $\boldsymbol{Y} := (Y_1, \ldots, Y_n)'$. For any $\tau \in \mathbb{T}$, where $\mathbb{T} := [t_0, t_1]$ is a parameter space for τ_0, consider the residual sum of squares

$$\begin{aligned} S_n(\alpha, \tau) &= n^{-1} \sum_{i=1}^{n} \left(Y_i - X_i'\beta - X_i'\delta \mathbf{1}\{Q_i < \tau\}\} \right)^2 \\ &= \|\boldsymbol{Y} - \boldsymbol{X}(\tau)\alpha\|_n^2, \end{aligned} \tag{1.2.3}$$

where $\alpha = (\beta', \delta')'$.

The scaled LASSO for threshold regression is defined as the one-step minimizer such that:

$$(\widehat{\alpha}, \widehat{\tau}) := \text{argmin}_{\alpha \in \mathscr{B} \subset \mathbb{R}^{2p}, \tau \in \mathbb{T} \subset \mathbb{R}} \{S_n(\alpha, \tau) + \lambda \|\boldsymbol{D}(\tau)\alpha\|_1\}, \tag{1.2.4}$$

where $\mathscr{B}$ is a parameter space for α_0 and λ is a tuning parameter chosen by the researcher which I discuss further in Section 1.3. The $(2p \times 2p)$ diagonal weighting matrix is denoted as follows:

$$\boldsymbol{D}(\tau) := \text{diag}\left\{ \left\| \boldsymbol{X}^{(j)}(\tau) \right\|_n, \;\; j = 1, ..., 2p \right\}, \tag{1.2.5}$$

where $\boldsymbol{X}^{(j)}(\tau)$ denotes the $j-th$ column of $\boldsymbol{X}(\tau)$. Furthermore, I can rewrite the ℓ_1 penalty as

$$\begin{aligned}\lambda\|\boldsymbol{D}(\tau)\alpha\|_1 &= \lambda\sum_{j=1}^{2p}\left\|\boldsymbol{X}^{(j)}(\tau)\right\|_n\left|\alpha^{(j)}\right| \\ &= \lambda\sum_{j=1}^{p}\left[\left\|X^{(j)}\right\|_n\left|\alpha^{(j)}\right| + \left\|X^{(j)}(\tau)\right\|_n\left|\alpha^{(p+j)}\right|\right],\end{aligned}$$

To be more exact, $(\widehat{\alpha},\widehat{\tau})$ in (1.2.4) can be regarded as a two-step minimizer such that:
Step 1.

For each $\tau\in\mathbb{T}$, $\widehat{\alpha}(\tau)$ is defined as

$$\widehat{\alpha}(\tau) := \operatorname{argmin}_{\alpha\in\mathscr{B}\subset\mathbb{R}^{2p}}\left\{S_n(\alpha,\tau)+\lambda\|\boldsymbol{D}(\tau)\alpha\|_1\right\}, \tag{1.2.6}$$

Step 2.

Define $\widehat{\tau}$ as the estimator of τ_0 such that:

$$\widehat{\tau} := \operatorname{argmin}_{\tau\in\mathbb{T}\subset\mathbb{R}}\left\{S_n(\widehat{\alpha}(\tau),\tau)+\lambda\|\boldsymbol{D}(\tau)\widehat{\alpha}(\tau)\|_1\right\}. \tag{1.2.7}$$

It is worth mentioning that $\widehat{\alpha}(\tau)$ is the weighted LASSO that uses a data-dependent ℓ_1 penalty to balance covariates adequately. Additionally, $\widehat{\tau}$ is an interval and in accordance with Lee et al. (2016), I define the maximum of the interval as the estimator $\widehat{\tau}$. For any n, it suffices in practice to search over $\{Q_1,\cdots Q_n\}$ as candidates for $\widehat{\tau}$, as these are the points where $\mathbf{1}\{Q_i>\tau\}, i=1,\cdots n$ will change. To put it another way, I think the parameter space $\mathbb{T}$ is divided into n intervals depending on $Q_1,\cdots Q_n$.

1.3 Oracle Inequalities

In this section, I establish the oracle inequality for the scaled LASSO estimator in (1.2.4). As I are considering a random design as opposed to a fixed regressor design in Lee et al. (2016), our assumptions are imposed in a slightly different form. Note Lee et al. (2016) have already argued how some of their assumptions could be valid in a random design.

Assumption 1. *Let $\{X_i,U_i,Q_i\}_{i=1}^n$ denote a sample of size n, where the covariates $\{X_i,Q_i\}_{i=1}^n$ are independently and identically distributed (i.i.d.). Additionally, the error terms $\{U_i\}_{i=1}^n$ are assumed to be independently distributed.*

(i)For the parameter space $\mathscr{B}$ for α_0, any $\alpha:=(\alpha_1,\cdots,\alpha_{2p})\in\mathscr{B}\subset\mathbb{R}^{2p}$,including α_0, satisfies $\|\alpha\|_\infty\le C_1$, for some constant $C_1>0$. $\mathscr{M}(\alpha_0)\le s_0$ and $\frac{s_0^2\|\delta_0\|_1^2\log p}{n}=o_p(1)$.

(ii) Marginal distribution of Q_i *is uniform* $(0,1)$. $\tau_0 \in \mathbb{T} = [t_0, t_1]$ *with* $0 < t_0 < t_1 < 1$.

(iii) $\max_{1\le j\le p} E\left[(X_i^{(j)})^4\right] \le C_2^4$ *and* $\min_{1\le j\le p} E\left[(X_i^{(j)}(t_0))^2\right] \ge C_3^2$ *uniformly in n for some universal constants* C_2 *and* C_3. $E\left[X_i^{(j)}X_i^{(l)}|Q_i=\tau\right]$ *is continuous and bounded when* τ *is in a neighborhood of* τ_0 *for all* $1 \le j, l \le p$.

(iv) The error terms $E(U_i|X_i,Q_i)=0$ *and* $E(U_i^2) \le C < \infty$ *for a positive universal constant C.*

(v) $\frac{\sqrt{EM_{UX}^2}\sqrt{\log p}}{\sqrt{n}} = o_p(1)$ *where* $M_{UX} = \max_{1\le i\le n}\max_{1\le j\le p}|U_iX_i^{(j)}|$.

(vi) $\frac{\sqrt{EM_{XX}^2}\sqrt{\log p}}{\sqrt{n}} = o_p(1)$ *where* $M_{XX} = \max_{1\le i\le n}\max_{1\le j,l\le p}|X_i^{(j)}X_i^{(l)} - E[X_i^{(j)}X_i^{(l)}]|$.

(vii) $\frac{\sqrt{EM_{Xt_0}^2}\sqrt{\log p}}{\sqrt{n}} = o_p(1)$, *where* $M_{Xt_0} = \max_{1\le i\le n}\max_{1\le j\le p}|(X_i^{(j)}(t_0))^2 - E(X_i^{(j)}(t_0))^2|$.

Assumption 1 states that the covariates are independently and identically distributed. The choice of identical distribution for the covariates is primarily motivated by maintaining simplicity in expressions, although there is flexibility to relax this assumption. Notably, the error terms are permitted to exhibit non-identical distribution, allowing for the possibility of conditional heteroskedasticity.

Assumption 1 (i) imposes restrictions for each component of the slope parameter vector. The second part of Assumption 1 (i) implies that s_0 and $\|\delta_0\|_1$ can increase with n.

Next, I describe how to solve the problem where the distribution of the threshold variable is not uniform. This technique is based on empirically transforming the distribution of the threshold variables to a uniform distribution. Suppose that the threshold variable $\{\tilde{Q}\}$ has a continuous distribution for which the cumulative distribution function is $F_{\tilde{Q}}$. The probability integral transform implies that the random variable Q has a standard uniform distribution where Q is defined as $Q = F_{\tilde{Q}}(\tilde{Q})$. To transform the marginals, I compute $Q_i = \widehat{F_{\tilde{Q}}}(\tilde{Q}_i) = \frac{\text{rank of } \tilde{Q}_i \text{ among } \{\tilde{Q}_i\}_{i=1}^n}{n}$, where $\widehat{F_{\tilde{Q}}}$ denotes the empirical distribution functions of the data $\left\{\tilde{Q}_i\right\}_{i=1}^n$. In particular, as a result of a continuous distribution, there is no tie among $\left\{\tilde{Q}_i\right\}_{i=1}^n$. We will show that the performance of our estimator does not depend on whether the threshold variable(Q_i) is part of the set of covariates(X_i) or correlated with the covariates in Section 1.5.

Assumption 1 (iii) to (vii) states restrictions on the covariates as well as the error terms in the random design setup studied in this article. Compared to Assumption 1 in Lee et al. (2016), I only assume the covariates and error terms are independently and identically distributed with uniformly bounded certain moments instead of sub-Gaussian data (Callot et al. (2017)) due to Chernozhukov et al. (2017). That is a much stronger assumption than the one imposed here and rules out data with heavy tails. Assumption 1 (ii) implies that $\min_{i=1,\cdots n} Q_i < t_0$. Intuitively, I assume that $\min_{1\le j\le p} E\left[(X_i^{(j)}(t_0))^2\right]$ is bounded away from 0.

Assumption 1 (iii) is a much stronger assumption than necessary conditions for the maximal inequality due to Chernozhukov et al. (2017). Apply the Cauchy-Schwarz Inequality to obtain the following: (i)$\max_{1\le j,l\le p} E\left[X_i^{(j)}X_i^{(l)}\right] \le C_2^2$ uniformly in n;

(ii)$\max_{1\le j\le p}\text{Var}(U_i X_i^{(j)})$, $\max_{1\le j\le p}\text{Var}(U_i X_i^{(j)}\mathbf{1}\{Q_i<\tau_0\})$, $\max_{1\le j,l\le p}\text{Var}(X_i^{(j)}X_i^{(l)})$, $\max_{1\le j,l\le p}\text{Var}(X_i^{(j)}X_i^{(l)}\mathbf{1}\{Q_i<\tau_0\})$, and $\max_{1\le j,l\le p}\text{Var}(X_i^{(j)}X_i^{(l)}\mathbf{1}\{Q_i<t_0\})$, are bounded away from infinity uniformly in n. Assumption 1 is used to establish the oracle inequality in Lemma 1, Theorem 1 and 2.

Now define

$$\lambda = \frac{C}{\mu}\frac{\sqrt{\log p}}{\sqrt{n}} \tag{1.3.1}$$

as the tuning parameter in (1.2.4) for a constant C and a fixed constant $\mu \in (0,1)$.

Lemma 1. *Under Assumption 1, let* $(\widehat{\alpha},\widehat{\tau})$ *be the LASSO estimator defined by* (1.2.4) *with* $\lambda = \frac{C}{\mu}\frac{\sqrt{\log p}}{\sqrt{n}}$ *for a constant* C *and a fixed constant* $\mu\in(0,1)$. *Then, with probability approaching 1*[1] *I have*

$$\left\|\widehat{f}-f_0\right\|_n \le \sqrt{(6+2\mu)C_1\sqrt{C_2^2+\mu\lambda}}\sqrt{s_0\lambda}. \tag{1.3.2}$$

Lemma 1 states that regardless of the linearity of the model, the prediction norm of the scaled LASSO estimator defined by (1.2.4) converges to 0, provided that $n\to\infty$, $p\to\infty$ and $s_0\lambda\to 0$. This, in turn, plays an important role for proving the oracle inequality in Theorem 1 for the case of linear models and Theorem 2 for nonlinear models.

Next, I turn towards the standard assumptions in high-dimensional regression models. To this end, define the population covariance matrix $\mathbf{\Sigma}(\tau) := E\left[\boldsymbol{X}_i(\tau)'\boldsymbol{X}_i(\tau)\right]$, $\boldsymbol{M} := E(X_i{}'X_i)$, $\boldsymbol{M}(\tau) := E\left[X_i(\tau)'X_i(\tau)\right]$ and $\boldsymbol{N}(\tau) := \boldsymbol{M}-\boldsymbol{M}(\tau)$. Then, $\mathbf{\Sigma}(\tau)$ can be represented by a 4-block matrix, i.e.

$$\mathbf{\Sigma}(\tau) = \begin{bmatrix} \boldsymbol{M} & \boldsymbol{M}(\tau) \\ \boldsymbol{M}(\tau) & \boldsymbol{M}(\tau) \end{bmatrix}.$$

The population uniform adaptive restricted eigenvalue is denoted by

$$\kappa(s_0,c_0,\mathbb{S},\mathbf{\Sigma}) = \min_{\tau\in\mathbb{S}} \min_{J_0\subset\{1,\dots,2p\},|J_0|\le s_0} \min_{\gamma\neq 0,\|\gamma_{J_0^c}\|_1\le c_0\sqrt{s_0}\|\gamma_{J_0}\|_2} \frac{(\gamma' E\left[\boldsymbol{X}_i(\tau)'\boldsymbol{X}_i(\tau)\right]\gamma)^{1/2}}{\|\gamma_{J_0}\|_2}.$$

or

$$\kappa(s_0,c_0,\mathbb{S},\boldsymbol{M}) = \min_{\tau\in\mathbb{S}} \min_{J_0\subset\{1,\dots,2p\},|J_0|\le s_0} \min_{\gamma\neq 0,\|\gamma_{J_0^c}\|_1\le c_0\sqrt{s_0}\|\gamma_{J_0}\|_2} \frac{(\gamma' E\left[X_i(\tau)'X_i(\tau)\right]\gamma)^{1/2}}{\|\gamma_{J_0}\|_2}.$$

---

[1] at least $1-\left(\frac{1}{p^{\tilde{C}_1}}+\tilde{C}_2\frac{EM_{X^2}^2}{n\log p}\right)-\left(\frac{1}{p^{\tilde{C}_3}}+\tilde{C}_4\frac{EM_{Xt_0}^2}{n\log p}\right)-\left(\frac{1}{p^{\tilde{C}_5}}+\tilde{C}_6\frac{EM_{UX}^2}{n\log p}\right)-\left(\frac{1}{(pn)^{\tilde{C}_7}}+\tilde{C}_8\frac{EM_{UX}^2}{n\log(pn)}\right)$, for some universal positive constants $\tilde{C}_1\cdots\tilde{C}_8$.

or

$$\kappa(s_0, c_0, \boldsymbol{M}) = \min_{J_0 \subset \{1,\ldots,2p\}, |J_0| \leq s_0} \quad \min_{\gamma \neq 0, \|\gamma_{J_0^c}\|_1 \leq c_0 \sqrt{s_0} \|\gamma_{J_0}\|_2} \frac{(\gamma' E\left[X_i{}' X_i\right] \gamma)^{1/2}}{\|\gamma_{J_0}\|_2}.$$

depending on the context.

In the literature on high-dimensional econometrics and statistics, it is common to add an adaptive restricted eigenvalue condition. Additionally, I consider that an adaptive restricted eigenvalue is of the same magnitude uniformly over $\tau \in \mathbb{T}$ as follows

Assumption 2. *(i) $\boldsymbol{M}$, $\boldsymbol{M}(\boldsymbol{\tau})$ and $\boldsymbol{N}(\boldsymbol{\tau})$ are are non-singular;*
(ii)[Uniform Adaptive Restricted Eigenvalue Condition] For some integer s_0 such that $\mathcal{M}(\alpha_0) \leq s_0 < p$, a positive number c_0 and some set $\mathbb{S} \subset \mathbb{R}$, the following condition holds

$$\kappa(s_0, c_0, \mathbb{S}, \boldsymbol{\Sigma}) > 0. \tag{1.3.3}$$

Assumption 2 (i) is a standard assumption in regression models. One can provide sufficient conditions for Assumption 2 (ii) by imposing the condition that the population covariance matrix $\boldsymbol{\Sigma}(\tau)$ have full rank. Hence, I are interested in property of $\boldsymbol{\Sigma}(\tau)$. To solve the inverse of the population covariance matrix, I do the Gaussian elimination to get

$$\boldsymbol{\Sigma}(\tau)^{-1} = \begin{bmatrix} \boldsymbol{N}(\boldsymbol{\tau})^{-1} & -\boldsymbol{N}(\boldsymbol{\tau})^{-1} \\ -\boldsymbol{N}(\boldsymbol{\tau})^{-1} & \boldsymbol{M}(\boldsymbol{\tau})^{-1} + \boldsymbol{N}(\boldsymbol{\tau})^{-1} \end{bmatrix}, \tag{1.3.4}$$

provided that $\boldsymbol{M}$, $\boldsymbol{M}(\boldsymbol{\tau})$ and $\boldsymbol{N}(\boldsymbol{\tau})$ are non-singular. Therefore, $\boldsymbol{\Sigma}(\tau)$ has full rank as long as $\boldsymbol{M}$, $\boldsymbol{M}(\boldsymbol{\tau})$ and $\boldsymbol{N}(\boldsymbol{\tau})$ are invertible. Thus, Assumption 2 (ii) is almost automatic under non-singularity conditions for $\boldsymbol{M}$, $\boldsymbol{M}(\boldsymbol{\tau})$ and $\boldsymbol{N}(\boldsymbol{\tau})$.

We will show that $\frac{1}{n}\boldsymbol{X}(\tau)'\boldsymbol{X}(\tau)$ uniformly converges to $\boldsymbol{\Sigma}(\tau)$ under Assumption 1 in Lemma 6 in the Appendix. Thus the empirical adaptive restricted eigenvalue condition can hold under the population eigenvalue condition imposed here, which can be seen in Lemma 7 in the Appendix.

Considering τ_0 is unknown, I impose that the restricted eigenvalue condition holds uniformly over τ. Intuitively, $\delta_0 \neq 0$ is a necessary condition of identifiability of τ_0. If $\delta_0 = 0$, I have to assume Assumption 2 holds with $\mathbb{S} = \mathbb{T}$, the whole parameter space for τ_0. By contrast, it suffices to impose the Adaptive Restricted Eigenvalue Condition holding uniformly on a neighborhood of τ_0, when $\delta_0 \neq 0$.

The Uniform Adaptive Restricted Eigenvalue Condition is crucial for us to update the boundness in Lemma 1. Lemma 1 states that the prediction norm is bounded by a factor of $s_0\lambda$. This bound is larger than what is desired for an oracle inequality. Depending on the UARE condition, the prediction norm as well as the ℓ_1 estimation error will be further tightened in the next section. Lee et al. (2016) proposed a type of slope estimator that is not affected by the presence of a threshold effect. That is to say, I can make predictions and estimate α_0 even if $\delta_0 = 0$ does not hold. However, I can derive oracle inequalities in terms of the prediction error and the ℓ_1 estimation error for unknown parameters α_0 separately in two cases depending on whether $\delta_0 = 0$ or not.

1.3.1 Case I. No Threshold.

First, I consider the situation where $\delta_0 = 0$. In this case, the true model is a linear model $Y_i = X_i'\beta_0 + U_i$, but I estimate it using the method defined by (1.2.4). Our estimated model is much more over-parametrized than the true one, but I shall obtain relatively precise estimates for the slope parameter vector α_0.

Theorem 1. *Supposed that $\delta_0 = 0$, let Assumptions 1-2 hold with $\kappa = \kappa(s_0, \frac{1+\mu}{1-\mu}, \mathbb{T}, \Sigma)$ for $0 < \mu < 1$. Let $(\widehat{\alpha}, \widehat{\tau})$ be the LASSO estimator defined by* (1.2.4) *with λ given by* (1.3.1)*. Then for all sufficiently large n, with probability approaching 1*[2] *I have*

$$\left\|\widehat{f} - f_0\right\|_n \leq \frac{2\sqrt{2}}{\kappa}\left(\sqrt{C_2^2 + \mu\lambda}\right)\sqrt{s_0}\lambda,$$
$$\|\widehat{\alpha} - \alpha_0\|_1 \leq \frac{4\sqrt{2}}{(1-\mu)\kappa^2}\frac{C_2^2 + \mu\lambda}{\sqrt{C_3^2 - \mu\lambda}}s_0\lambda.$$

Furthermore, these bounds are valid uniformly over the l_0-ball

$$\mathscr{A}_{\ell_0}^{(1)}(s_0) = \left\{\alpha_0 \in \mathbb{R}^{2p} \mid \|\beta_0\|_\infty \leq C_1, \mathscr{M}\left(\beta_0\right) \leq s_0, \delta_0 = 0\right\}.$$

It is worth noting that the bound of the prediction norm here is much smaller than in Lemma 1. Compared to Theorem 2 in Lee et al. (2016) or the oracle inequality in the literature on high-dimensional linear models (Bickel et al. (2009), van de Geer et al. (2014) etc.), Theorem 1 delivers results of the same magnitude. Although our model is much more overparametrized, our estimation method can accommodate the linear model. Nonetheless, there is a variable selection problem on δ_0. Our estimation method can find more nonzero coefficients than the true number. In particular,, some $\delta(\widehat{\tau})_j$ is incorrectly estimated as nonzero. We shall discuss this in more detail in Section A.8.

[2] at least $1 - \left(\frac{1}{p^{\tilde{C}_1}} + \tilde{C}_2\frac{EM_{X2}^2}{n\log p}\right) - \left(\frac{1}{p^{\tilde{C}_3}} + \tilde{C}_4\frac{EM_{Xt_0}^2}{n\log p}\right) - \left(\frac{1}{p^{\tilde{C}_5}} + \tilde{C}_6\frac{EM_{UX}^2}{n\log p}\right) - \left(\frac{1}{(pn)^{\tilde{C}_7}} + \tilde{C}_8\frac{EM_{UX}^2}{n\log(pn)}\right) - \left(\frac{1}{p^{2\tilde{C}_9}} + \tilde{C}_{10}\frac{EM_{XX}^2}{n\log p^2}\right) - \left(\frac{1}{(p^2n)^{\tilde{C}_{11}}} + \tilde{C}_{12}\frac{EM_{XX}^2}{n\log(p^2n)}\right)$, for some universal positive constants $\tilde{C}_1 \cdots \tilde{C}_{12}$.

1.3.2 Case II. Fixed Threshold.

In this subsection, I construct oracle inequalities when $\delta_0 \neq 0$. More explicitly, the true model has a well-identified and discontinuous threshold effect.

Assumption 3 (Identifiability under Sparsity and Discontinuity of Regression)**.** *For a given* $s_0 \geq \mathcal{M}(\alpha_0)$, *and for any* η *and* τ *such that* $\eta < |\tau - \tau_0|$ *and* $\alpha \in \{\alpha : \mathcal{M}(\alpha) \leq s_0\}$, *there exists a constant* $C_4 > 0$ *such that wpa1*

$$\left\| f_{(\alpha,\tau)} - f_0 \right\|_n^2 > C_4 \eta.$$

Assumption 3 states identifiability of τ_0. Lee et al. (2016) have already discussed in Appendix B.1. (page. A7–A8) that Assumption 3 is valid in a random design under Assumption 1 above. As mentioned before, I need Assumption 2 to hold uniformly in a neighborhood of τ_0. Lemma 9 shows how I can get an upper bound of $\hat{\tau} - \tau_0$ only under Assumption 1 and 3.

Given Lemma 9 in the appendix, I define

$$\eta^* = \frac{2(3+\mu)C_1}{C_4} \sqrt{C_2^2 + \mu\lambda s_0}\lambda$$

and

$$\mathbb{S} = \left\{ |\tau - \tau_0| \leq \eta^* \right\},$$

where $\mathbb{S}$ can be inserted into Assumption 2. Note that I omit the restriction, $\eta \geq \min_i |Q_i - \tau_0|$ which is imposed in Lee et al. (2016). The reason is that $\eta \geq \min_i |Q_i - \tau_0|$ never binds for sufficiently large n. Intuitively, $\min_i |Q_i - \tau_0|$ will be small enough in the random design.

Assumption 4 (Smoothness of Design)**.** *For any* $\eta > 0$, *there exists a constant* $C_5 < \infty$ *such that wpa1*

$$\sup_{1 \leq j,l \leq p} \sup_{|\tau - \tau_0| < \eta} \frac{1}{n} \sum_{i=1}^{n} \left| X_i^{(j)} X_i^{(l)} \right| |\mathbf{1}\{Q_i < \tau_0\} - \mathbf{1}\{Q_i < \tau\}| \leq C_5 \eta, \tag{1.3.5}$$

$$\sup_{1 \leq j,l \leq p} \sup_{|\tau - \tau_0| < \eta} \|\delta_0\|_1 \left| \frac{1}{n} \sum_{i=1}^{n} U_i X_i^{(j)} [\mathbf{1}\{Q_i < \tau_0\} - \mathbf{1}\{Q_i < \tau\}] \right| \leq \frac{\lambda\sqrt{\eta}}{2}, \tag{1.3.6}$$

$$\sup_{|\tau - \tau_0| < \eta} \left| \frac{1}{n} \sum_{i=1}^{n} U_i X_i' \delta_0 [\mathbf{1}\{Q_i < \tau_0\} - \mathbf{1}\{Q_i < \tau\}] \right| \leq \frac{\lambda\sqrt{\eta}}{2}. \tag{1.3.7}$$

Lemme 4 demonstrates that $\sup_{1 \leq j \leq p} \frac{1}{n} \sum_{i=1}^{n} U_i X_i^{(j)}$ is bounded by λ with probability approaching one (wpa1). Similarly, Lemma 6 shows that $\sup_{1 \leq j,l \leq p} |\frac{1}{n} \sum_{i=1}^{n} X_i^{(j)} X_i^{(l)}|$ is bounded from above wpa1. The supremum in Assumption 4 is bounded in a neighborhood of τ_0 for all $1 \leq j, l \leq p$. This strengthening is essential to establish oracle inequalities when a threshold is present. (1.3.5) is the case when the Q_i are continuously distributed and $E\left(X_i^{(j)}X_i^{(l)}|Q_I=\tau\right)$ is continuous and bounded in a neighborhood of τ_0 for all $1 \geq j,l \geq p$. It is worth noting that (1.3.7) almost automatically implies (1.3.6). Furthermore, Lemma 10 demonstrates that if a design satisfies Assumption 1, then Assumption 4 holds.

Theorem 2. *Suppose that $\delta_0 \neq 0$, let Assumption1 to 2 hold with $\kappa = \kappa(s_0, \frac{2+\mu}{1-\mu}, \mathbb{S}, \boldsymbol{\Sigma})$ for $0<\mu<1$. Furthermore, Assumptions 3 and 4 hold. Let $(\widehat{\alpha}, \widehat{\tau})$ be the LASSO estimator defined by* (1.2.4) *with λ given by* (1.3.1).

Then for all sufficiently large n, with probability approaching 1[3] *I have*

$$\left\|\widehat{f}-f_0\right\|_n \leq 6\frac{\sqrt{C_2^2+\mu\lambda}}{\kappa}\sqrt{s_0}\lambda,$$

$$\|\widehat{\alpha}-\alpha_0\|_1 \leq \frac{36(C_2^2+\mu\lambda)}{\kappa^2(1-\mu)\sqrt{C_3^2-\mu\lambda}}s_0\lambda,$$

$$|\widehat{\tau}-\tau_0| \leq \left(\frac{3\left(1+\mu\right)\sqrt{\left(C_2^2+\mu\lambda\right)}}{(1-\mu)\sqrt{\left(C_3^2-\mu\lambda\right)}}+1\right)\frac{12(C_2^2+\mu\lambda)}{\kappa^2 C_4}s_0\lambda^2.$$

Furthermore, these bounds are valid uniformly over the l_0-ball

$$\mathscr{A}_{\ell_0}^{(2)}(s_0) = \left\{\alpha_0 \in \mathbb{R}^{2p} \mid \|\alpha_0\|_\infty \leq C_1, \mathscr{M}(\alpha_0) \leq s_0, \delta_0 \neq 0\right\}.$$

The oracle inequalities in Theorem 2, disregarding constant terms, align with those in Theorem 1 concerning the prediction norm and $\ell 1$ errors for estimates. These results hold uniformly over $\mathscr{B}_{\ell_0}(s_0)$, where

$$\mathscr{B}_{\ell_0}(s_0) = \mathscr{A}_{\ell_0}^{(2)}(s_0) \cup \mathscr{A}_{\ell_0}^{(1)}(s_0) = \left\{\alpha_0 \in \mathbb{R}^{2p} \mid \|\alpha_0\|_\infty \leq C, \mathscr{M}(\alpha_0) \leq s_0\right\}.$$

For the super-consistency result of $\widehat{\tau}$, Lee et al. (2016) argued that the least squares objective function behaving locally linearly around the true threshold parameter value is the key to achieving the super-consistency for the threshold parameter.

The main contribution of this section is that I have extended the oracle inequality to non-sub-Gaussian random regressors with non-sub-Gaussian errors for both the prediction norm and ℓ_1 errors for estimates.

[3] at least $1-\left(\frac{1}{p^{\tilde{C}_1}}+\tilde{C}_2\frac{EM_{X2}^2}{n\log p}\right)-\left(\frac{1}{p^{\tilde{C}_3}}+\tilde{C}_4\frac{EM_{Xt_0}^2}{n\log p}\right)-\left(\frac{1}{p^{\tilde{C}_5}}+\tilde{C}_6\frac{EM_{UX}^2}{n\log p}\right)-\left(\frac{1}{(pn)^{\tilde{C}_7}}+\tilde{C}_8\frac{EM_{UX}^2}{n\log(pn)}\right)-\left(\frac{1}{p^{2\tilde{C}_9}}+\tilde{C}_{10}\frac{EM_{XX}^2}{n\log p^2}\right)-\left(\frac{1}{(p^2n)^{\tilde{C}_{11}}}+\tilde{C}_{12}\frac{EM_{XX}^2}{n\log(p^2n)}\right)$, for some universal positive constants $\tilde{C}_1\cdots\tilde{C}_{12}$.

1.4 The Debiased LASSO

Now, I turn to the construction of confidence bands for the elements of α_0, ensuring uniform validity over all α_0 within certain ℓ_0-balls. To achieve this, I introduce the debiased LASSO estimator, which is employed in the construction of tests and confidence intervals. Specifically, I consider the following form of the debiased LASSO estimator:

$$\widehat{a}(\widehat{\tau}) = \widehat{\alpha}(\widehat{\tau}) + \widehat{\Theta}(\widehat{\tau})\boldsymbol{X}'(\widehat{\tau})(Y - \boldsymbol{X}(\widehat{\tau})\widehat{\alpha}(\widehat{\tau}))/n, \tag{1.4.1}$$

where $\widehat{\alpha}(\widehat{\tau})$ and $\widehat{\tau}$ is defined in (1.2.4), $\widehat{\Theta}(\widehat{\tau})$ is an approximate inverse of the Gram matrix $\widehat{\Sigma}(\widehat{\tau}) = \boldsymbol{X}'(\widehat{\tau})\boldsymbol{X}(\widehat{\tau})/n$. The reason that I calculate an approximate inverse of the Gram matrix to be used is that when $p > \widehat{\tau}n$,[4], an inverse of the Gram matrix $\widehat{\Sigma}(\widehat{\tau})$ is not feasible. $\widehat{\Sigma}(\tau)$ is of reduced rank provided that $p > \widehat{\tau}n$. Thus, the idea is to construct an approximate inverse, $\widehat{\Theta}(\widehat{\tau})$, to $\widehat{\Sigma}(\widehat{\tau})$ and control the error term resulting from this approximation. Our construction of this approximate inverse relies on nodewise regression of Yuan (2010), which will be introduced in the next subsection.

We now derive the error decomposition of the estimator in (1.4.1), which provides intuitions on the construction of $\widehat{\Theta}(\widehat{\tau})$.

By the minimizing property of $\widehat{\alpha}(\widehat{\tau})$, it follows the first-order condition of (1.2.6):

$$-\boldsymbol{X}(\widehat{\tau})'(Y - \boldsymbol{X}(\widehat{\tau})\widehat{\alpha}(\widehat{\tau}))/n + \lambda \boldsymbol{D}(\widehat{\tau})\widehat{\boldsymbol{\rho}} = 0, \tag{1.4.2}$$

where $\widehat{\rho}$ is a $2p$ by 1 vector, arising from the subdifferential of $\|\widehat{\alpha}(\widehat{\tau})\|_1$. $\|\widehat{\rho}\|_\infty \leq 1$ and $\widehat{\rho}_j = \text{sign}(\widehat{a}^{(j)}(\widehat{\tau}))$ if $\widehat{a}^{(j)}(\widehat{\tau}) \neq 0$, where "sign()" is the function that maps positive entries to 1 and negative entries to -1. So (1.4.1) can be rewritten as

$$\widehat{a}(\widehat{\tau}) = \widehat{\alpha}(\widehat{\tau}) + \lambda \widehat{\Theta}(\widehat{\tau})\boldsymbol{D}(\widehat{\tau})\widehat{\boldsymbol{\rho}} \tag{1.4.3}$$

1.4.1 Bias Correction Case I. No Threshold

We first consider the case where $\delta_0 = 0$. Then, the true model is simply a linear model $Y_i = X_i'\beta_0 + U_i$.

[4]Given that τ_0 is unknown, I must construct the approximate inverse $\widehat{\Theta}(\tau)$ when $\widehat{\tau} \cdot n < p$. In a more stringent assumption, if $\min\{t_0, t_1, 1-t_0, 1-t_1\} \cdot n < p$, I construct the approximate inverse $\widehat{\Theta}(\tau)$.

Inserting $Y = X\beta_0 + U$ into (1.4.2) yields

$$\lambda \boldsymbol{D}(\widehat{\tau})\widehat{\rho} + \boldsymbol{X}(\widehat{\tau})'(\boldsymbol{X}(\widehat{\tau})\widehat{\alpha}(\widehat{\tau}) - X\beta_0)/n = \boldsymbol{X}'(\widehat{\tau})U/n. \tag{1.4.4}$$

Note that $\delta_0 = 0$ implies $X(\widehat{\tau})\delta_0 = 0$ for any $\widehat{\tau} \in \mathbb{T}$, since $\widehat{\tau}$ is an overparameterized term, I can add $X(\widehat{\tau})\delta_0$ into (1.4.4) anywhere to obtain

$$\lambda \boldsymbol{D}(\widehat{\tau})\widehat{\rho} + \boldsymbol{X}(\widehat{\tau})'(\boldsymbol{X}(\widehat{\tau})\widehat{\alpha}(\widehat{\tau}) - \boldsymbol{X}(\widehat{\tau})\alpha_0)/n = \boldsymbol{X}'(\widehat{\tau})U/n, \tag{1.4.5}$$

The expression with one more step:

$$\lambda \boldsymbol{D}(\widehat{\tau})\widehat{\rho} + \widehat{\Sigma}(\widehat{\tau})(\widehat{\alpha}(\widehat{\tau}) - \alpha_0) = \boldsymbol{X}'(\widehat{\tau})U/n. \tag{1.4.6}$$

Thus, I have

$$\begin{aligned} \widehat{a}(\widehat{\tau}) &= \widehat{\alpha}(\widehat{\tau}) + \lambda\widehat{\Theta}(\widehat{\tau})\boldsymbol{D}(\widehat{\tau})\widehat{\rho} \\ &= \widehat{\alpha}(\widehat{\tau}) - \widehat{\Theta}(\widehat{\tau})\widehat{\Sigma}(\widehat{\tau})(\widehat{\alpha}(\widehat{\tau}) - \alpha_0) + \widehat{\Theta}(\widehat{\tau})\boldsymbol{X}'(\widehat{\tau})U/n \\ &= \alpha_0 - \alpha_0 + \widehat{\alpha}(\widehat{\tau}) - \widehat{\Theta}(\widehat{\tau})\widehat{\Sigma}(\widehat{\tau})(\widehat{\alpha}(\widehat{\tau}) - \alpha_0) + \widehat{\Theta}(\widehat{\tau})\boldsymbol{X}'(\widehat{\tau})U/n \\ &= \alpha_0 + \widehat{\Theta}(\widehat{\tau})\boldsymbol{X}'(\widehat{\tau})U/n - \Delta(\widehat{\tau})/n^{1/2}, \end{aligned} \tag{1.4.7}$$

where

$$\Delta(\tau) = \sqrt{n}(\widehat{\Theta}(\tau)\widehat{\Sigma}(\tau) - I_{2p})(\widehat{\alpha}(\tau) - \alpha_0).$$

In order to derive the asymptotic distribution of tests involving an increasing number of parameters, I define a $(2p \times 1)$ vector g with $\|g\|_2 = 1$ and let $H = j = 1, \cdots 2p \mid g_j \neq 0$ with cardinality $|H| = h < p$. H contains the indices of the coefficients involved. This implies $\|g\|_1 \leq \sqrt{h}$ by the Cauchy–Schwarz inequality. In particular, $g = e_j$ is the case where I only consider a single coefficient, where e_j is the $j-th$ $2p \times 1$ unit vector.

Considering

$$\sqrt{n}g'(\widehat{a}(\widehat{\tau}) - \alpha_0) = g'\widehat{\Theta}(\widehat{\tau})\boldsymbol{X}'(\widehat{\tau})U/n^{1/2} - g'\Delta(\widehat{\tau}), \tag{1.4.8}$$

a central limit theorem for $g'\widehat{\Theta}(\widehat{\tau})\boldsymbol{X}'(\widehat{\tau})U/n^{1/2}$ and a verification of the asymptotic negligibility of $g'\Delta(\widehat{\tau})$ will achieve the desired convergence and yield asymptotically Gaussian inference.

1.4.2 Bias Correction Case II. Fixed Threshold.

This subsection explores the case where the threshold effect is well identified and discontinuous. Following a similar derivation in Section 1.4.1, this time I insert $Y = \boldsymbol{X}(\tau_0)\alpha_0 + U$ into (1.4.2). This yields:

$$\lambda \boldsymbol{D}(\widehat{\tau})\widehat{\rho} + \widehat{\Sigma}(\widehat{\tau})\widehat{\alpha}(\widehat{\tau}) - \boldsymbol{X}'(\widehat{\tau})\boldsymbol{X}(\tau_0)\alpha_0/n = \boldsymbol{X}'(\widehat{\tau})U/n. \tag{1.4.9}$$

substituting (1.4.9) into (1.4.3) yields:

$$\begin{aligned} \widehat{\alpha}(\widehat{\tau}) = \alpha_0 + \widehat{\Theta}(\widehat{\tau})\boldsymbol{X}'(\widehat{\tau})(\boldsymbol{X}(\tau_0)\alpha_0 - \boldsymbol{X}(\widehat{\tau})\alpha_0)/n \\ - \widehat{\Theta}(\widehat{\tau})\lambda \boldsymbol{D}(\widehat{\tau})\widehat{\rho} + \widehat{\Theta}(\widehat{\tau})\boldsymbol{X}'(\widehat{\tau})U/n - \Delta(\widehat{\tau})/n^{1/2}. \end{aligned} \tag{1.4.10}$$

In order to derive the asymptotic distribution of tests involving an increasing number of parameters, I define a $(2p \times 1)$ vector g with $\|g\|_2 = 1$ and let $H = \{j = 1, \cdots 2p \mid g_j \neq 0\}$ with cardinality $|H| = h < p$. H contains the indices of the coefficients involved. This implies $\|g\|_1 \leq \sqrt{h}$ by Cauchy–Schwarz inequality. In particular, $g = e_j$ is the case where I only consider a single coefficient, where e_j is the $j-th$ $2p \times 1$ unit vector.

Considering

$$\begin{aligned} \sqrt{n} g'(\widehat{\alpha}(\widehat{\tau}) - \alpha_0) = g'\widehat{\Theta}(\tau_0)\boldsymbol{X}'(\tau_0)U//n^{1/2} - g'\Delta(\widehat{\tau}) \\ + g'(\widehat{\Theta}(\widehat{\tau})\boldsymbol{X}'(\widehat{\tau})U - \widehat{\Theta}(\tau_0)\boldsymbol{X}'(\tau_0)U)/n^{1/2} \\ + g'\widehat{\Theta}(\widehat{\tau})(\boldsymbol{X}'(\widehat{\tau})\boldsymbol{X}(\tau_0) - \boldsymbol{X}'(\widehat{\tau})\boldsymbol{X}(\widehat{\tau}))\alpha_0/n^{1/2}, \end{aligned} \tag{1.4.11}$$

a central limit theorem for $g'\widehat{\Theta}(\tau_0)\boldsymbol{X}'(\tau_0)U//n^{1/2}$ and a verification of asymptotic negligibility of $g'(\widehat{\Theta}(\widehat{\tau})\boldsymbol{X}'(\widehat{\tau})U - \widehat{\Theta}(\tau_0)\boldsymbol{X}'(\tau_0)U)//n^{1/2}$, $g'\Delta(\widehat{\tau})$ and $g'\widehat{\Theta}(\widehat{\tau})(\boldsymbol{X}'(\widehat{\tau})\boldsymbol{X}(\tau_0) - \boldsymbol{X}'(\widehat{\tau})\boldsymbol{X}(\widehat{\tau}))\alpha_0/n^{1/2}$ will achieve the desired convergence and yield asymptotically Gaussian inference.

1.4.3 Constructing the Approximate Inverse $\widehat{\Theta}(\tau)$

In this section, I formalize the approximate inverse $\widehat{\Theta}(\tau)$ utilized in our threshold model. The approach closely follows that of van de Geer et al. (2014), with the additional requirement of verifying that our specified conditions are met.

For the purpose discussed in the context of (1.4.11), I seek a well-behaved $\widehat{\Theta}(\tau)$ and examine the asymptotic properties of $\widehat{\Theta}(\tau)$ uniformly across $\tau \in \mathbb{T}$. To achieve this, I establish a connection between $\widehat{\Theta}(\tau)$ and $\Theta(\tau)$, where $\Theta(\tau) := \Sigma(\tau)^{-1}$ is defined in (1.3.4) as

$$\mathbf{\Sigma}(\tau)^{-1} = \begin{bmatrix} \boldsymbol{N}(\boldsymbol{\tau})^{-1} & -\boldsymbol{N}(\boldsymbol{\tau})^{-1} \\ -\boldsymbol{N}(\boldsymbol{\tau})^{-1} & \boldsymbol{M}(\boldsymbol{\tau})^{-1} + \boldsymbol{N}(\boldsymbol{\tau})^{-1} \end{bmatrix}.$$

Define $\widehat{\boldsymbol{M}}(\tau) = \frac{1}{n}\sum_{i=1}^{n} X_i{}'X_i\mathbf{1}\{Q_i < \tau\}$ and $\widehat{\boldsymbol{N}}(\tau) = \frac{1}{n}\sum_{i=1}^{n} X_i{}'X_i\mathbf{1}\{Q_i \geq \tau\}$. We construct the approximate inverse $\widehat{\Theta}(\tau)$ because $p > \widehat{\tau}n$.To be precise, the threshold variable Q_i is used to split the sample into two groups. As long as either sample covariance matrix $\widehat{\boldsymbol{M}}(\tau)$ or $\widehat{\boldsymbol{N}}(\tau)$ is of reduced rank, I have to construct their respective approximate inverses.

Then I construct approximate inverse $\widehat{\boldsymbol{A}}(\tau)$ of $\widehat{\boldsymbol{M}}(\tau)$ and $\widehat{\boldsymbol{B}}(\tau)$ of $\widehat{\boldsymbol{N}}(\tau)$ and I relate $\widehat{\boldsymbol{A}}(\tau)$ to $\boldsymbol{A}(\tau) := \boldsymbol{M}(\tau)^{-1}$ and $\widehat{\boldsymbol{B}}(\tau)$ to $\boldsymbol{B}(\tau) := \boldsymbol{N}(\tau)^{-1}$.

Let $X^{(-j)}(\tau)$ denote all columns of $X(\tau)$ except for the $j-th$ one and let $\tilde{X}^{(j)}(\tau)$ denote the $(n\times 1)$ vector such that $\tilde{X}_i^{(j)}(\tau) = X_i^{(j)}\mathbf{1}\{Q_i \geq \tau\}$. Then, $\tilde{X}^{(-j)}(\tau)$ denotes a $(n\times(p-1))$ matrix except for the $j-th$ column of $\tilde{X}(\tau)$. Along Section 2.1 of Yuan (2010) I can rewrite the following regression models with covariates orthogonal in L_2 to the error terms for all $j = 1\cdots p$,

$$X^{(j)}(\tau) = X^{(-j)}(\tau)'\gamma_{0,j}(\tau) + \upsilon^{(j)},$$

$$\tilde{X}^{(j)}(\tau) = \tilde{X}^{(-j)}(\tau)'\tilde{\gamma}_{0,j}(\tau) + \tilde{\upsilon}^{(j)}.$$

The details of the covariance matrix's representation of the regression coefficients are given in Appendix B of Caner and Kock (2018). $\upsilon^{(j)}$ and $\tilde{\upsilon}^{(j)}$ are not a function of τ due to the independence of Q.

We put forward the following assumptions:

Assumption 5. *(i) For the parameter space* $\max_{1\leq j\leq p}\|\gamma_j\|_\infty \leq C$ *, for some constant* $C > 0$*;*
(ii)$E(\upsilon_i^{(j)}|X_i, Q_i) = 0$ *and* $E[(\upsilon_i^{(j)})^2]$ *is uniformly bounded over* $j = 1,\cdots p$*;* $E(\tilde{\upsilon}_i^{(j)}|X_i, Q_i) = 0$ *and* $E[(\tilde{\upsilon}_i^{(j)})^2]$ *is uniformly bounded over* $j = 1,\cdots p$*;*
(iii) $\frac{\sqrt{EM_{\upsilon X}^2}\sqrt{\log p}}{\sqrt{n}} < \infty$*, where* $M_{\upsilon X} = \max_{1\leq i\leq n}\max_{1\leq l\leq p}|\upsilon_i^{(j)}X_i^{(l)}|$.

Assumption 5 controls the tail distribution of $|\upsilon_i^{(j)}X_i^{(l)}|$ and $|\tilde{\upsilon}_i^{(j)}X_i^{(l)}|$, in order to apply the oracle inequality proved in previous work.

Given any $\tau \in \mathbb{T}$, the LASSO nodewise regression for $\widehat{\boldsymbol{A}}(\tau)$ is defined as follows:

(1.4.12) $$\widehat{\gamma}_j(\tau) = \text{argmin}_{\gamma\in\mathbb{R}^{p-1}}\|X^{(j)}(\tau) - X^{(-j)}(\tau)\gamma\|_n^2 + \lambda_{node}\|\widehat{\mathbf{\Gamma}}_j(\tau)\gamma\|_1,$$

where

$$\widehat{\mathbf{\Gamma}}_j(\tau) := \text{diag}\left\{\left\|\boldsymbol{X}^{(l)}(\tau)\right\|_n, l = 1,\ldots,p, l \neq j\right\},$$

with components of $\widehat{\gamma}_j(\tau) = \{\widehat{\gamma}_j^{(k)}(\tau);\ k = 1,\ldots,p,\ k \neq j\}$. The $(2p\times 2p)$ diagonal weighting matrix is denoted as follows: It is noteworthy that I choose λ to be the same in all of the nodewise regressions. The nodewise LASSO runs p times as an intermediate step to construct $\widehat{\boldsymbol{A}}(\tau)$. Let

$$\widehat{C}(\tau) = \begin{pmatrix} 1 & -\widehat{\gamma_1}^{(2)}(\tau) & \cdots & -\widehat{\gamma_1}^{(p)}(\tau) \\ -\widehat{\gamma_2}^{(1)}(\tau) & 1 & \cdots & -\widehat{\gamma_2}^{(p)}(\tau) \\ \cdots & \cdots & \ddots & \cdots \\ -\widehat{\gamma_p}^{(1)}(\tau) & -\widehat{\gamma_p}^{(2)}(\tau) & \cdots & 1 \end{pmatrix}. \tag{1.4.13}$$

Then, model (1.4.12) will be sparse with $\widehat{\gamma_j}(\tau)$ possessing $s_j(\tau)$ non-zero entries. To define $\widehat{\boldsymbol{A}}(\tau)$ I introduce a $p \times p$ diagonal matrix $\widehat{Z}(\tau)^2 = diag(\widehat{z_1}(\tau)^2, \cdots, \widehat{z_p}(\tau)^2)$, where

$$\widehat{z_j}(\tau)^2 = \|X^{(j)}(\tau) - X^{(-j)}(\tau)\widehat{\gamma}_j(\tau)\|_n^2 + \lambda_{node}\|\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{j}}(\tau)\widehat{\gamma}_j(\tau)\|_1, \tag{1.4.14}$$

for all $j = 1, ..., p$. Hence, I may define

$$\widehat{\boldsymbol{A}}(\tau) = \widehat{Z}(\tau)^{-2}\widehat{C}(\tau). \tag{1.4.15}$$

It remains to be shown that this $\widehat{\boldsymbol{A}}(\tau)$ is close to the inverse of $\widehat{\boldsymbol{M}}(\tau)$. We define $\widehat{A}_j(\tau)$ as the $j-th$ row of $\widehat{\boldsymbol{A}}(\tau)$. Thus, $\widehat{A}_j(\tau) = \widehat{C_j}(\tau)/\widehat{z_j}(\tau)^2$. Denoting by $\tilde{e}_j$ the $j-th$ $p \times 1$ unit vector, the KKT conditions also imply that

$$\|\widehat{A}_j(\tau)'\widehat{\boldsymbol{M}}(\tau) - \tilde{e}_j'\|_\infty \le \frac{\lambda_{node}}{\widehat{z_j}(\tau)^2}. \tag{1.4.16}$$

Parallel to construction of $\widehat{\boldsymbol{A}}(\tau)$ above, I define

$$\widehat{\boldsymbol{B}}(\tau) = \widehat{\tilde{Z}}(\tau)^{-2}\widehat{\tilde{C}}(\tau). \tag{1.4.17}$$

We define

$$\widehat{\tilde{\gamma}}(\tau)_j = \text{argmin}_{\gamma\in\mathbb{R}^{p-1}}\|\tilde{X}^{(j)}(\tau) - \tilde{X}^{(-j)}(\tau)'\gamma\|_n^2 + \lambda_{\text{node}}\|\widehat{\tilde{\boldsymbol{\Gamma}}}_{\boldsymbol{j}}(\tau)\gamma\|_1,$$
$$\widehat{\tilde{\boldsymbol{\Gamma}}}_{\boldsymbol{j}}(\tau) = \text{diag}\left\{\left\|\tilde{\boldsymbol{X}}^{(l)}(\tau)\right\|_n,\ l = 1, ..., p,\ l \ne j\right\},$$

with components of $\widehat{\tilde{\gamma}}(\tau)_j = \{\widehat{\tilde{\gamma}_j}^{(k)}(\tau) : k = 1, ..., p, k \ne j\}$.

Denote by $\widehat{\tilde{Z}}(\tau)^2 = \text{diag}(\widehat{\tilde{z}_1}(\tau)^2, \cdots, \widehat{\tilde{z}_p}(\tau)^2)$, which is a $p \times p$ diagonal matrix with

$$\widehat{\tilde{z}}_j(\tau)^2 = \|\tilde{X}^{(j)}(\tau) - \tilde{X}^{(-j)}(\tau)'\widehat{\tilde{\gamma}}(\tau)_j\|_n^2 + \lambda_{\text{node}}\|\widehat{\tilde{\boldsymbol{\Gamma}}}_{\boldsymbol{j}}(\tau)\widehat{\tilde{\gamma}}(\tau)_j\|_1,$$

for all $j = 1, ..., p$.

We let

$$\widehat{C}(\tau) = \begin{pmatrix} 1 & -\widehat{\gamma}_1^{(2)}(\tau) & \cdots & -\widehat{\gamma}_1^{(p)}(\tau) \\ -\widehat{\gamma}_2^{(1)}(\tau) & 1 & \cdots & -\widehat{\gamma}_2^{(p)}(\tau) \\ \cdots & \cdots & \ddots & \cdots \\ -\widehat{\gamma}_p^{(1)}(\tau) & -\widehat{\gamma}_p^{(2)}(\tau) & \cdots & 1 \end{pmatrix}. \tag{1.4.18}$$

We also get the following inequality:

$$\|\widehat{B}(\tau)'_j \widehat{\boldsymbol{N}}(\tau) - \tilde{e}'_j\|_\infty \leq \frac{\lambda_{node}}{\widehat{\tilde{z}}_j(\tau)^2}. \tag{1.4.19}$$

Thus

$$\widehat{\boldsymbol{\Theta}}(\tau) = \begin{bmatrix} \widehat{\boldsymbol{B}}(\tau) & -\widehat{\boldsymbol{B}}(\tau) \\ -\widehat{\boldsymbol{B}}(\tau) & \widehat{\boldsymbol{A}}(\tau) + \widehat{\boldsymbol{B}}(\tau) \end{bmatrix}. \tag{1.4.20}$$

Denoting by e_j the $j-th$ $2p \times 1$ unit vector,

$$\max_{j \in H} \sup_{\tau \in \mathbb{T}} \|\widehat{\boldsymbol{\Theta}}(\tau)'_j \widehat{\boldsymbol{\Sigma}}(\tau) - e'_j\|_\infty \leq \max_{j \in H \text{or} j+p \in H} \sup_{\tau \in \mathbb{T}} \frac{\lambda_{node}}{\widehat{z_j}(\tau)^2} + \max_{j \in H \text{or} j+p \in H} \sup_{\tau \in \mathbb{T}} \frac{\lambda_{node}}{\widehat{\tilde{z}}_j(\tau)^2}. \tag{1.4.21}$$

(Formal proof is given in the Appendix.)

Hence, I get the error term resulting from this approximation, i.e. the upper bound on the maximal absolute entry of the $j-th$ row of $\widehat{\boldsymbol{\Theta}}(\tau)'\widehat{\boldsymbol{\Sigma}}(\tau) - \boldsymbol{I}_{2p}$. This provides the sufficient conditions to show that $g'\Delta(\tau)$ in (1.4.11) is asymptotically negligible.

Define $\bar{s} = \sup_{\tau \in \mathbb{T}} \max_{j \in H} s_j(\tau)$, where $s_j(\tau) = |S_j(\tau)|$, and$S_j(\tau) = \{\Theta_{j,i}(\tau) \neq 0\}$.

We then have the following result.

Lemma 2. *Let Assumptions 1-5 be satisfied and set* $\lambda_{node} = \frac{C}{\mu}\sqrt{\frac{\log p}{n}}$. *Then,*

$$\max_{j \in H} \sup_{\tau \in \mathbb{T}} \|\widehat{\boldsymbol{\Theta}}(\tau)_j - \boldsymbol{\Theta}(\tau)_j\|_1 = O_p\left(\bar{s}\sqrt{\frac{\log p}{n}}\right) \tag{1.4.22}$$

$$\max_{j \in H} \sup_{\tau \in \mathbb{T}} \|\widehat{\boldsymbol{\Theta}}(\tau)_j - \boldsymbol{\Theta}(\tau)_j\|_2 = O_p\left(\sqrt{\frac{\bar{s}\log p}{n}}\right) \tag{1.4.23}$$

$$\max_{j \in H} \sup_{\tau \in \mathbb{T}} \|\widehat{\boldsymbol{\Theta}}(\tau)_j\|_1 = O_p\left(\sqrt{\bar{s}}\right) \tag{1.4.24}$$

$$\max_{j \in H} \sup_{\tau \in \mathbb{T}} \|\widehat{\boldsymbol{\Theta}}(\tau)'_j \widehat{\boldsymbol{\Sigma}}(\tau) - e'_j\|_\infty = O_p\left(\sqrt{\frac{\log p}{n}}\right) \tag{1.4.25}$$

1.4.4 Inference

In this section, I derive asymptotic normality under high-level conditions which allows us to establish joint inference on a linear combination of the entries of the debiased LASSO $\widehat{a}(\widehat{\tau})$.

To this end, I define

$$\widehat{\Sigma}_{xu}(\widehat{\tau}) = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i(\widehat{\tau})'\boldsymbol{X}_i(\widehat{\tau})(\widehat{U}_i(\widehat{\tau}))^2,$$

$$\tilde{\kappa}(s_0, c_0, \mathbb{T}, \Sigma) = \max_{\tau\in\mathbb{T}} \max_{J_0\subset\{1,\ldots,2p\},|J_0|\le s_0} \max_{\gamma\neq 0,\|\gamma_{J_0^c}\|_1\le c_0\sqrt{s_0}\|\gamma_{J_0}\|_2} \frac{(\gamma' E\left[\boldsymbol{X}_i(\tau)'\boldsymbol{X}_i(\tau)\right]\gamma)^{1/2}}{\|\gamma_{J_0}\|_2},$$

and

$$\widehat{\tilde{\kappa}}(s_0, c_0, \mathbb{T}, \widehat{\Sigma}) = \max_{\tau\in\mathbb{T}} \max_{J_0\subset\{1,\ldots,2p\},|J_0|\le s_0} \max_{\gamma\neq 0,\|\gamma_{J_0^c}\|_1\le c_0\sqrt{s_0}\|\gamma_{J_0}\|_2} \frac{(\gamma' \frac{1}{n}\boldsymbol{X}(\tau)'\boldsymbol{X}(\tau)\gamma)^{1/2}}{\|\gamma_{J_0}\|_2}.$$

The following assumptions are imposed to establish a limiting distribution for an increasing number of coefficients.

Assumption 6. *(i)* $\max_{1\le j\le p} E\left[(X_i^{(j)})^{12}\right]$ *and* $E\left[U_i^4\right]$ *are bounded away from infinity uniformly in n.*
$\frac{\sqrt{EM_{X^6}^2}\sqrt{\log p}}{\sqrt{n}} = o_p(1)$, $\frac{\sqrt{EM_{X^2U^2}^2}\sqrt{\log p}}{\sqrt{n}} = o_p(1)$ *and* $\frac{\sqrt{EM_{X^4U^2}^2}\sqrt{\log p}}{\sqrt{n}} = o_p(1)$, *where*

$$M_{X^6} = \max_{1\le i\le n}\max_{1\le k,l,j\le p} |(X_i^{(k)}X_i^{(l)}X_i^{(j)})^2 - E(X_i^{(k)}X_i^{(l)}X_i^{(j)})^2|,$$

$$M_{X^2U^2} = \max_{1\le i\le n}\max_{1\le j,l\le p} |X_i^{(j)}X_i^{(l)}U_i^2 - E[X_i^{(j)}X_i^{(l)}U_i^2]|,$$

and

$$M_{X^4U^2} = \max_{1\le i\le n}\max_{1\le j,l\le p} |(X_i^{(j)}X_i^{(l)}U_i)^2 - E[X_i^{(j)}X_i^{(l)}U_i]^2|.$$

(ii)

$$(h)^{\frac{3}{2}} s_0^2 \bar{s}^2 \frac{\log p}{\sqrt{n}} = o_p(1).$$

(iii) $\frac{(h\bar{s})^3}{n^2} = o_p(1)$;
$\kappa(\bar{s}, c_0, \mathbb{T}, \Sigma_{xu})$ *and* $\kappa(\bar{s}, c_0, \mathbb{T}, \Sigma)$ *are bounded away from zero;*
$\tilde{\kappa}(\bar{s}, c_0, \mathbb{T}, \Sigma_{xu})$ *and* $\tilde{\kappa}(\bar{s}, c_0, \mathbb{T}, \Sigma)$ *are bounded from above.*

Assumption 6 gives sufficient conditions for a central limit theorem result. Assumption 6 (i) controls the tail behavior of the covariates and the error terms. By Assumption 6(i), $\max_{1\le j,l\le p}\mathrm{Var}(X_i^{(j)}X_i^{(l)}U_i^2)$, $\max_{1\le k,l,j\le p}\mathrm{Var}(X_i^{(k)}X_i^{(l)}X_i^{(j)})^2$ and $\max_{1\le j,l\le p}\mathrm{Var}(X_i^{(j)}X_i^{(l)}U_i)^2$ are bounded away from infinity uniformly in n.

Assumption 6(ii) limits the dimension of the models, the dimension involved in conducting joint inference, the sparsity of the population covariance matrix, and the sparsity of the slope parameter vector.

The first part of Assumption 6 (iii) is designed to verify the Lyapunov condition. Then the other part restricts the eigenvalues of $\Sigma_{xu}(\tau)$ and $\Sigma(\tau)$.

Hence, I have the following result.

Theorem 3. *Let Assumptions 1, 2, 3, 4, 5 and 6 be satisfied and let g be $2p\times 1$ vector satisfying $\|g\|_2 = 1$. Then,*

$$\frac{\sqrt{n}g'(\widehat{a}(\widehat{\tau})-\alpha_0)}{\sqrt{g'\widehat{\Theta}(\widehat{\tau})\widehat{\Sigma}(\widehat{\tau})_{xu}\widehat{\Theta}(\widehat{\tau})'g}} \xrightarrow{d} N(0,1). \tag{1.4.26}$$

Furthermore,

$$\sup_{\alpha_0\in\mathscr{A}^{(1)}_{\ell_0}(s_0)} |g'\widehat{\Theta}(\widehat{\tau})\widehat{\Sigma}_{xu}(\widehat{\tau})\widehat{\Theta}'(\widehat{\tau})g - g'\Theta(\widehat{\tau})\Sigma_{xu}(\widehat{\tau})\Theta'(\widehat{\tau})g| = o_p(1) \tag{1.4.27}$$

$$\sup_{\alpha_0\in\mathscr{A}^{(2)}_{\ell_0}(s_0)} \sup_{\tau_0\in\mathbb{T}} |g'\widehat{\Theta}(\widehat{\tau})\widehat{\Sigma}_{xu}(\widehat{\tau})\widehat{\Theta}'(\widehat{\tau})g - g'\Theta(\tau_0)\Sigma_{xu}(\tau_0)\Theta'(\tau_0)g| = o_p(1) \tag{1.4.28}$$

where

$$\mathscr{A}^{(1)}_{\ell_0}(s_0) = \left\{\alpha_0 \in \mathbb{R}^{2p} \mid \|\alpha_0\|_\infty \le C, \mathscr{M}(\alpha_0) \le s_0, \delta_0 = 0\right\},$$

and

$$\mathscr{A}^{(2)}_{\ell_0}(s_0) = \left\{\alpha_0 \in \mathbb{R}^{2p} \mid \|\alpha_0\|_\infty \le C, \mathscr{M}(\alpha_0) \le s_0, \delta_0 \ne 0\right\}.$$

When conducting testing, I lack prior knowledge of the presence of a threshold. Consequently, in Theorem 3, I need to simultaneously impose the assumptions of Theorems 1 and 2. The first part of Theorem 3 implies convergence to the normal distribution of a sub-vector of $\widehat{a}(\widehat{\tau})$ of increasing dimension uniformly over $\mathscr{B}_{\ell_0}(s_0)$. The number of parameters involved in hypotheses is allowed to grow to infinity at a rate restricted by the above Assumption 6(ii).

The second part shows that I propose a consistent estimator of the covariance matrix uniformly over $\mathscr{B}_{\ell_0}(s_0)$. The uniformity of (1.4.27) and (1.4.28) will also be used in the proof of uniform convergence below.

In the case where H is a set of fixed cardinality h, (1.4.26) implies that

$$\|\left(g'\widehat{\Theta}(\widehat{\tau})\widehat{\Sigma}(\widehat{\tau})_{xu}\widehat{\Theta}(\widehat{\tau})'g\right)^{-\frac{1}{2}}\sqrt{n}g'(\widehat{a}(\widehat{\tau})-\alpha_0)\|_2^2 \xrightarrow{d} \chi^2(h), \tag{1.4.29}$$

correspondingly $\|g\|_2 = 1$ and $H = \{j = 1, \cdots 2p \mid g_j \ne 0\}$. Thus, χ^2 test can be conducted with a hypothesis on h parameters simultaneously.

We now establish confidence intervals for our parameters. We refer to the proof of Theorem 3 in Caner and Kock (2018) and its details therefore are omitted.

Let $\Phi(t)$ denote the cumulative distribution function (CDF) of the standard normal distribution and $z_{1-\frac{\alpha}{2}}$ is the $1-\frac{\alpha}{2}$ percentile of the standard normal distribution. Denote by $\widehat{\sigma}(\widehat{\tau})_j = \sqrt{e_j'\widehat{\Theta}(\widehat{\tau})\widehat{\Sigma}(\widehat{\tau})_{xu}\widehat{\Theta}(\widehat{\tau})'e_j}$ for all $j \in \{1, \cdots 2p\}$. Let diam($[a,b]$) map the length of the interval$[a,b] \subset \mathbb{R}$.

Hence I have the following results:

Theorem 4. *Let Assumptions 1, 2, 3, 4, 5 and 6 be satisfied and let* g *be* $2p \times 1$ *vector satisfying* $\|g\|_2 = 1$. *Then,*

$$\sup_{t\in\mathbb{R}} \sup_{\alpha_0 \in \mathcal{A}_{\ell_0}(s_0)} \left| \mathbb{P}\left\{ \frac{\sqrt{n}g'(\widehat{a}(\widehat{\tau}) - \alpha_0)}{\sqrt{g'\widehat{\Theta}(\widehat{\tau})\widehat{\Sigma}(\widehat{\tau})_{xu}\widehat{\Theta}(\widehat{\tau})'g}} \leq t \right\} - \Phi(t) \right| \to 0. \tag{1.4.30}$$

Furthermore, for all $j \in \{1, \cdots 2p\}$,

$$\lim_{n\to\infty} \inf_{\alpha_0 \in \mathcal{B}_{\ell_0}(s_0)} \mathbb{P}\left\{ \alpha_0^{(j)} \in \left[\widehat{a}^{(j)}(\widehat{\tau}) - z_{1-\frac{\alpha}{2}} \frac{\widehat{\sigma}(\widehat{\tau})_j}{\sqrt{n}}, \widehat{a}^{(j)}(\widehat{\tau}) + z_{1-\frac{\alpha}{2}} \frac{\widehat{\sigma}(\widehat{\tau})_j}{\sqrt{n}} \right] \right\} = 1 - \alpha. \tag{1.4.31}$$

Finally,

$$\sup_{\alpha_0 \in \mathcal{B}_{\ell_0}(s_0)} diam(\left[\widehat{a}^{(j)}(\widehat{\tau}) - z_{1-\frac{\alpha}{2}} \frac{\widehat{\sigma}(\widehat{\tau})_j}{\sqrt{n}}, \widehat{a}^{(j)}(\widehat{\tau}) + z_{1-\frac{\alpha}{2}} \frac{\widehat{\sigma}(\widehat{\tau})_j}{\sqrt{n}} \right]) = O_p(\frac{1}{\sqrt{n}}), \tag{1.4.32}$$

where

$$\mathcal{B}_{\ell_0}(s_0) = \mathcal{A}_{\ell_0}^{(2)}(s_0) \cup \mathcal{A}_{\ell_0}^{(1)}(s_0) = \left\{ \alpha_0 \in \mathbb{R}^{2p} \mid \|\alpha_0\|_\infty \leq C, \mathcal{M}(\alpha_0) \leq s_0 \right\}.$$

(1.4.30) implies that the convergence to the standard normal distribution is actually valid uniformly over the ℓ_0-ball of radius at most s_0.

1.5 Monte Carlo Simulation

In this section, I explore the finite sample properties of the proposed debiased LASSO procedures for threshold regression through Monte Carlo experiments. To establish a benchmark for the debiased LASSO in threshold regression, I also implement the debiased LASSO for linear regression as introduced by van de Geer et al. (2014). Before delving into the results, I provide an explanation of how the data was generated and the performance measures used for comparing the debiased LASSO for the threshold model with the debiased LASSO method proposed by van de Geer et al. (2014).

1.5.1 Implementation Details

The implementation of the debiased LASSO for linear model is inspired by the publicly available code at `https://web.stanford.edu/~montanar/ssLASSO/code.html`. We also modify the code of Callot et al. (2017) at `https://github.com/lcallot/ttlas` to the debiased LASSO for threshold model. To choose the tuning parameters λ_{node}, I employ the Generalized Information Criterion proposed by Konishi and Kitagawa (1996) (GIC). We utilize GIC and ten-fold cross-validation to select the tuning parameters λ. However, according to our simulation results, cross-validation does not significantly enhance the quality of the results, while the processing time is considerably longer. Hence, I select both λ and λ_{node} based on GIC.

1.5.2 Performance Measures

We focus successively on several dimensions: the number of observations, the quantity of covariates, and the correlation between the threshold variable and the covariates.

Both covariates and error terms are assumed to follow a t-distribution with 10 degrees of freedom. Specifically, each covariate is generated as $X_i^{(j)} \sim t(10)$ for each $j \in \{1,\ldots,p\}$, and the error term $U_i \sim t(10)$ is independent of the covariates. When the threshold variable Q_i is independent of X_i, $Q_i \sim \text{uniform}(0,1)$. In the case where the threshold variable Q_i is correlated with X_i, Q_{1i} and Q_{2i} are generated from uniform$(0,1)$ distributions, and ρ_i is generated from a Bernoulli distribution with a success probability equal to the correlation.Subsequently, I generate $Q_i = \rho_i Q_{1i} + (1-\rho_i)Q_{2i}$; then, I replace the second column of X_i with Q_{1i}, and the correlation between Q_i and $X_i^{(2)}$ equals the success probability. We set the threshold parameter $\tau_0 = 0.5$ to prevent either of the split samples from being unbalanced. Neither the intercept nor the thresholded intercept is involved in the design to simplify the test. Without loss of generality, I assume that β_0 is a $p \times 1$ vector with the first $s_0/2$ elements being ones and the remaining $p - s_0/2$ elements being zeros. Additionally, I assume that the sparsity patterns of both β_0 and δ_0 are identical, resulting in a total of s_0 nonzero parameters.

All tests are conducted at a 5% significance level, and all confidence intervals are set at the 95% level. The χ^2-test involves the first non-zero parameter and the first zero parameter in β_0 and δ_0 for threshold regression. For linear regression, the χ^2-test involves the first non-zero parameter and the first zero parameter in β_0.

The performance of our debiased LASSO for threshold regression and the debiased LASSO of van de Geer et al. (2014) is evaluated based on the following statistics, averaged across iterations.

1. Size: We evaluate the size of the χ^2-test in (1.4.29) for a hypothesis involving more than one parameter. The null hypothesis is always that the coefficients equal the true value.

2. Power: We evaluate the size of the χ^2-test in (1.4.29) for a hypothesis involving more than one parameter. To measure the power of the test, I test whether $\delta_0^{(s_0/2+1)}$ equals its assigned value plus 1. The difference in alternatives is merely to obtain non-trivial power comparisons (i.e. to avoid either the power of all tests being (very close to) zero or (very close to) one).

3. Coverage Rate: We compute the coverage rate of a Gaussian confidence interval constructed as in (1.4.32) in Theorem 4. All results related to the debiased LASSO for threshold regression are conducted for the coefficients corresponding to the first nonzero entry and first zero entry of β_0 (i.e., $\beta_0^{(1)}, \beta_0^{(s_0/2+1)}$) and the first nonzero entry and first zero entry of δ_0 (i.e., $\delta_0^{(1)}, \delta_0^{(s_0/2+1)}$). All results related to the debiased LASSO for linear regression are conducted for the coefficients corresponding to the first nonzero entry and first zero entry of β_0 (i.e., $\beta_0^{(1)}, \beta_0^{(s_0/2+1)}$).

For assessing the size of the χ^2-test using the debiased LASSO for threshold regression, I evaluate the true hypothesis $H_0 : (\beta_0^{(1)}, \beta_0^{(s_0/2+1)}, \delta_0^{(1)}, \delta_0^{(s_0/2+1)}) = (1,0,1,0)$. Due to an incorrect model specification, the χ^2-test for the debiased LASSO for linear regression focuses on the first nonzero entry and first zero entry of β_0 i.e., $H_0 : (\beta_0^{(1)}, \beta_0^{(s_0/2+1)}) = (1,0)$.

To evaluate the power of the χ^2-test, I examine the false hypothesis $H_0 :$ $(\beta_0^{(1)}, \beta_0^{(s_0/2+1)}, \delta_0^{(1)}, \delta_0^{(s_0/2+1)}) = (1,0,1,1)$ for the debiased LASSO for threshold regression. For the debiased LASSO for linear regression, I test the false hypothesis $H_0 : (\beta_0^{(1)}, \beta_0^{(s_0/2+1)}) = (1,1)$.

We construct confidence intervals for $(\beta_0^{(1)}, \beta_0^{(s_0/2+1)}, \delta_0^{(1)}, \delta_0^{(s_0/2+1)})$ for the debiased LASSO for threshold regression or $(\beta_0^{(1)}, \beta_0^{(s_0/2+1)})$ for the debiased LASSO for linear regression.

The number of Monte Carlo replications for each design is consistently set to 200.

1.5.3 Design 1

In this design, I investigate the effect of using a threshold variable that is part of the set of covariates ($Q \in X$), or that is correlated with the covariate. To be more precise, let $X^{(2)}$ denote the second column of X, and $\rho_{Q,X^{(2)}}$ be the correlation between Q and $X^{(2)}$. We consider the case where Q is independent of X, $Q = X^{(2)}$, as well as $\rho_{Q,X^{(2)}} = 0.9$.

1.5.4 Design 2

This design is to increase the number of observations or the number of variables to investigate the asymptotic properties of our procedure. We take $s_0 = 10$, which is a very sparse setting to satisfy assumptions on s_0 with relatively large n and p.

The following combinations of n and p are considered:

$$(n, p) \in \{(500,100),(500,250),(500,400),(100,250),(300,250)\}$$

1.5.5 Results of Simulations

In this section, I present the results of a series of simulation experiments assessing the finite sample properties of the debiased LASSO for threshold regression.

Table 1.1 indicates that whether the threshold variable is included in the set of covariates or correlated with one of the covariates has almost no impact on the performance of our debiased Lasso Estimator for the high-dimensional threshold model. The size and power of the χ^2-test are very close to the nominal significance level with the debiased LASSO for the threshold model. Notably, it is not surprising that the size and power of the χ^2-test exhibit substantial distortion with the debiased LASSO for the linear model. All numbers concerning confidence bands are reasonable with the debiased LASSO for the threshold model but are slightly over-covered. The debiased Lasso Estimator for the linear model procedure does exhibit an undercover coverage rate of nonzero parameters, but for zero parameters, it has a tendency to overcover.

Table 1.2 indicates that the debiased LASSO for the high-dimensional threshold model consistently exhibits less size distortion while having slightly more power as n increases in a high-dimensional setting. The size and power approach nominal levels as n is increased. All numbers concerning coverage rates are reasonable with the debiased LASSO for the threshold model. The debiased LASSO for the threshold model procedure always has coverages that gradually improve as the sample size is increased. However, nonzero coverage rates are close to 1, indicating a tendency to overcover.

Table 1.3 illustrates that the debiased LASSO for the threshold model procedure continues to perform well in terms of size, power, and the confidence intervals even as experiments become progressively more challenging, with the choice of $p > n$ and the models naturally becoming high-dimensional. However, nonzero coverage rates are close to 1, indicating a tendency to overcover.

In general, the debiased LASSO for threshold regression performs much better in terms of size, power, and coverage rate compared to the debiased LASSO for the linear model proposed by van de Geer et al. (2014) when threshold effects are present.

1.6 Conclusion

In this paper, I introduce a debiased LASSO estimation procedure designed for high-dimensional threshold models. We propose a method for constructing uniformly valid confidence bands in the context of the nonlinear regime-switch model. Notably, our study adopts less restrictive assumptions compared to existing research on high-dimensional threshold mod-

Table 1.1: Summary Statistics for Design 1: the Dpendence Between the Threshold Variable and the Regressors

		χ^2 size	χ^2 power	coverage rate non-zero β	coverage rate non-zero δ	coverage rate zero β	coverage rate zero δ
		$n=500, p=250$					
Q is independent of X	DLTH	0.03	0.93	0.99	0.98	1.00	1.00
	DL	0.43	0.57	0.27		1.00	
$\rho_{Q,X^{(2)}} = 0.9$	DLTH	0.05	0.86	0.95	0.98	1.00	1.00
	DL	0.49	0.50	0.23		1.00	
$Q = X^{(2)}$	DLTH	0.04	0.91	0.97	0.98	1.00	1.00
	DL	0.44	0.46	0.21		1.00	

els, removing the assumption of sub-Gaussian error terms and covariates. Future research directions could include extending the debiased LASSO methodology to dynamic panels with threshold effects and expanding the framework to accommodate models with multiple thresholds.

Table 1.2: Summary Statistics for Design 2: the Number of Observations

		χ^2		coverage rate			
		$p=250$, $Q=X^{(2)}$					
		size	power	non-zero		zero	
				β	δ	β	δ
n=100	DLTH	0.12	0.19	0.90	0.77	1.00	1.00
	DL	0.03	0.26	0.96		1.00	
n=300	DLTH	0.09	0.72	0.93	0.94	1.00	1.00
	DL	0.21	0.34	0.78		1.00	
n=500	DLTH	0.04	0.91	0.97	0.98	1.00	1.00
	DL	0.44	0.46	0.21		1.00	

Table 1.3: Summary Statistics for Design 2: The Number of Variables

		χ^2		coverage rate			
		n=500, $Q=X^{(2)}$					
		size	power	non-zero		zero	
				β	δ	β	δ
p=100	DLTH	0.05	0.97	0.95	0.95	1.00	1.00
	DL	0.75	0.65	0.08		1.00	
p=250	DLTH	0.04	0.91	0.97	0.98	1.00	1.00
	DL	0.44	0.46	0.21		1.00	
p=400	DLTH	0.02	0.92	0.99	0.97	1.00	1.00
	DL	0.34	0.44	0.41		1.00	

CHAPTER

2

INVESTIGATING INTEGRATION AND EXCHANGE RATE PASS-THROUGH IN WORLD MAIZE MARKETS USING DEBIASED LASSO INFERENCE

2.1 Introduction

Efficient markets are expected to eliminate any potential for riskless profits through arbitrage and trade, a result known as the "Law of One Price" (LOP). Economic arbitrage relies on the principle that prices of related goods should not arbitrarily differ from one another over the long run. The general implication here is that prices for homogeneous products at different geographic locations in otherwise freely functioning markets should differ by no more than transport and transactions costs. However, the existence of transactions costs can introduce a threshold effect, where deviations in prices above a certain threshold are necessary to trigger price movements. In recent years, studies analyzing this phenomenon have focused on developing nonlinear models that can better capture the effects of unobservable transaction costs in spatial price linkages. The motivation behind using such models is to better understand the dynamics of market integration and the role of transaction costs in the presence of regime changes. The use of nonlinear models has been largely driven by the application of threshold modeling techniques. These models are based on the idea that transaction costs and other barriers to spatial trade may lead to regime switching, with alternative regimes representing the trade and no-trade equilibria. This idea has been operationalized through various econometric techniques and model specifications.

Threshold autoregression (TAR) models have indeed had a significant impact on the analysis of asymmetric price transmission in agricultural economics. These models have been developed to capture the nonlinear dynamics of market integration and account for the effects of unobserved transaction costs that can affect spatial price linkages. A common approach to threshold modeling often involves an autoregressive model of the price differential. The study conducted by Goodwin and Piggott (2001) examined corn prices at local markets by combining a threshold structure with an error-correction model. Goodwin et al. (1990) noted that delivery lags that extend beyond a single time period may imply arbitrage conditions that involve noncontemporaneous price linkages. Based on this idea, Lence et al. (2018) examined the performance of the threshold cointegration approach, specifically Band-TVECM, in analyzing price transmission in an explicit context where trade decisions are made based on the expectation of final prices because trade takes time. In addition to the threshold model, Goodwin et al. (2021) applied generalized additive models to empirical considerations of price transmission and spatial market integration.

The extensive literature addressing price transmission and market integration has largely focused on simple comparisons of prices in geographically distinct markets. In rare cases when the consideration applies to international market comparisons, exchange rates may also be considered. It is, however, possible that a number of other variables may be relevant to market linkages. These variables may largely proxy for unobservable transactions costs or other reasons for price differences, such as government policies and product heterogeneity. The puzzle confronting the analyst is which of these variables makes valid and useful proxies for the unobservable reasons as to why prices may differ. We attempt to address this puzzle by considering data-driven methods for selecting the optimal set of variables relevant to an understanding of price linkages in international maize markets. Specifically, we apply a high dimensional threshold model to examine the effect of exchange rates and market factors on price linkages among spatially distinct world maize markets. Such an application is a natural methodological extension of existing empirical studies on spatial market integration models.

Although exchange-rate pass-through, i.e. the degree to which exchange rate movements are reflected in prices has long been a question of interest in international economics, there is limited literature that examines exchange-rate pass-through in global agricultural commodity markets. One study by Chambers and Just (1981) uses an econometric model of the wheat, corn, and soybean markets to investigate the dynamic effects of exchange rate fluctuations on U.S. commodity markets. The study finds that exchange rate fluctuations have a significant real impact on agricultural markets, particularly on the volume of exports and the relative split between exports and domestic use of these commodities. The econometric model developed in the study shows that agricultural prices are sensitive to movements in the exchange rate, with short-run adjustments being more dramatic than longer-run adjustments. Varangis and Duncan (1993) employed a system of equations, including one for the Japanese export price and another for the US producer price, to estimate the impact of changes in the yen/dollar exchange rate and other factors on steel prices. It is worth noting that this trade is unidirectional, as Japanese steel is consistently imported into the USA. The study demonstrates that such exchange rate fluctuations are not fully passed through to steel prices. The pricing decisions of US steel producers are primarily influenced by changes in their production costs and the US index of industrial production, rather than fluctuations in the yen/dollar exchange rate.

International trade in basic commodities is generally invoiced in US dollar terms. At first glance, this may seem to imply that exchange rates are irrelevant to market linkages. However, assuming that the commodities are valued in local currencies after being imported suggests that exchange rates may still be relevant to price linkages. We discuss this point in greater detail below.

Barrett and Li (2002) examine actual trade flows as a factor for assessing spatial market integration. They note that empirical tests should differentiate between the notions of spatial market integration and a competitive market equilibrium. The latter concept refers to market conditions where no trade occurs because arbitrage conditions do not provide opportunities for profitable trading. The authors highlight that prices in two segmented markets might react to exogenous factors like inflation or climatic conditions without representing a spatial equilibrium in markets. A recent overview from the World Bank Rebello (2020) addresses the factors influencing spatial market integration. The overview mentions the cooperation among policymakers on matters such as trade and investment policies, migration, transportation infrastructure, macroeconomic policy, natural resource policy, and others related to "shared sovereignty." Furthermore, the overview highlights the critical role of regional integration in policy reforms, contributing significantly to overall peace and security.

The integration of world markets for grains and oilseeds has been of interest for many years. In recent years, the global maize market has been dominated by major exporters such as the United States, Argentina, and Ukraine, which have consistently ranked among the top maize producers and exporters worldwide. The US, the largest producer, alone accounts for over one-third of global maize exports. Argentina and Ukraine collectively account for over onefourth of global maize exports. The dominance of these countries in the global maize market is representative of the market and makes them candidates for studying price transmission and market integration. They play a crucial role in global maize prices and influencing maize markets worldwide. Likewise, the extent to which distortions arise due to incomplete pass-through of exchange rate shocks has been an important indicator of the overall functions of markets.

In addition to prices and exchange rates, other market factors can be conceptually related to market linkages, such as aggregate economic indicators like industrial production, trade policies, and exogenous shocks, such as the recent pandemic, interest rates, and nominal inflation rates in each market. These factors may be associated with deviations from perfect market integration, as they can affect the costs of transportation, communication, and transactions between markets, as well as the demand and supply conditions in each market. Understanding the effects of these market factors on price linkages is essential for policymakers and market participants to make informed decisions about trade, investment, and risk management.

LASSO (least absolute shrinkage and selection operator) is a regression technique that uses shrinkage methods for variable selection. LASSO employs L1 regularization and shrinkage techniques to penalize the model based on the absolute value of parameter estimates. It is a valid approach for identifying an optimal model specification by selecting the variables that contribute the most to explaining a regression-type relationship. Although LASSO models have been widely used in economics studies, the shrinkage bias introduced due to the penalization in the LASSO loss function can affect the properly scaled limiting distribution of the LASSO estimator. Therefore, to conduct valid statistical inference, we need to remove this bias. This paper uses the desparsified (debaised) LASSO (least absolute shrinkage and selection operator) method for high dimensional threshold regression, recently developed by Yan (2023) to model the nonlinearity in the spatial price integration models. The fact is that existing literature on price transmission and exchange rate pass-through has developed from simple regression models to nonlinear specifications that allow differential impacts on price linkages. These differential effects are often identified using smooth or discrete threshold models.

2.2 Econometrics Models of Spatial Market Integration

Spatial market integration in agricultural product markets has been extensively studied in the literature. Consider a commodity traded in common currency in two regional or international markets represented by location indices j and k. The individual market prices are denoted by P^j and P^k, respectively. The arbitrage condition of perfect market integration reflects the equation $P_t^j/P_t^k = 1$, abstracting from trade and transportation costs. This condition has been adjusted to account for the wedge between prices due to transaction or transportation costs, which may differ significantly in regional markets. The general representation for this adjusted arbitrage condition is $1/(1-\kappa) \leq P_t^j/P_t^k \leq 1-\kappa$, where κ represents the proportional loss in commodity value due to transaction or transportation costs ($0 < \kappa < 1$). The greater the distance between locations j and k, the closer κ is to one. It should be noted that many factors may be relevant to price differences across markets. Most existing studies have only considered simple price relationships. An important distinction exists between transportation and transactions costs, which include transport costs as well as other factors that contribute to price differences. These factors could include variables associated with economic and trade policies, product characteristics, and risk.

Many spatial economic models utilize the iceberg trade cost proposed by Samuelson (1954), which assumes that part of the produced output representing the material costs of transportation melts away during transportation. That is, after taking natural logarithms and denoting $p_t^j = \ln P_t^j$, the inequality is often presented as

$$|p_t^j - p_t^k| \leq \ln(1-\kappa). \tag{2.2.1}$$

The inequality (2.2.1) is generally considered to reflect two distinct states of the market. The first state corresponds to a condition where there is no profitable trading, with $|p_t^1 - p_t^2| \leq \ln(1-\kappa)$. Under conditions of trade or profitable arbitrage opportunities, the condition holds as $|p_t^j - p_t^k| > \ln(1-\kappa)$. The speed at which the market adjusts to such deviations from the arbitrage equilibrium is often used as a measure of the degree of market integration. Typically, these discrete arbitrage and no-arbitrage conditions are represented using threshold models, where the threshold represents an empirical measure of the transaction cost, $\ln(1-\kappa)$. Bidirectional trade models may allow for different thresholds depending on which market price is higher.

Over time, log price differentials within the band limits are expected to follow a unit root process. Conversely, log price differences outside the band are expected to be mean-reverting, which suggests the existence of a transactions cost band, as assumed in the literature.

A wide literature has examined spatial market integration in world markets for agricultural commodities. Likewise, a large related literature has examined how shocks to exchange rates affect domestic and export prices, a phenomenon known as 'pass-through'. If a shock to exchange rates is fully reflected in adjustments to prices, the shock is considered to have been fully passed through. Most empirical studies of market integration and exchange rate pass-through assume a linear relationship, as represented by

$$p_t^j = \alpha + \beta p_t^k + \gamma_2 \pi_t^{jk} + \varepsilon_t, \tag{2.2.2}$$

where p_t^j is the price in market j in time period t and π_t^{jk} is the exchange rate between currencies in markets j and k, all in logarithmic terms.

Perfect integration is implied if $\alpha = 0$ and $\beta = 1$. In cases where prices are invoiced in different currencies, perfect integration also requires perfect exchange rate pass-through, which is implied if $\gamma_2 = 1$. If prices are invoiced in a common currency, as is often the case when trade is conducted in US dollar terms, the exchange rate is 1 and thus the logarithmic value of zero eliminates the exchange rate effect[1]. However, exchange rate distortions may still affect price linkages, which is implied if $\gamma_2 \neq 0$, even if prices are quoted in a common currency. This could occur if imported goods are moved into internal markets in which a different currency applies.

It is also essential to consider other market factors associated with deviations from perfect integration. To this end, we consider an alternative version of equation (2.2.2) that is expressed as:

$$p_t^j - p_t^k = \gamma_2 \pi_t^{jk} + \gamma_3 Z_t^{jk} + \varepsilon_t, \tag{2.2.3}$$

where Z_t^{jk} is a set of factors that may be conceptually relevant to price linkage, γ_3 is a vector of parameters corresponding to Z_t^{jk}. These factors include exogenous shocks such as exchange rates, interest rates, unemployment rates, and nominal inflation rates in each of the markets. We do not know which, if any of these factors, is likely to be relevant to price linkages. These factors largely proxy for unobservable factors that may be related to price relationships, such as local policies, product heterogeneity, and unobservable transactions costs.

To further analyze spatial price linkages, we evaluate deviations from a price parity condition, considering threshold effects of price differentials and isolated shocks in spatially distinct markets. In addition to the conventional specification, we propose an extension to this framework of spatial market integration that includes two regimes. One regime represents a case of no trade, while another represents conditions of profitable trade and arbitrage. The regime switch depends on a forcing variable, usually a lagged price differential, expressed as:

$$\begin{aligned} \Delta(p_t^j - p_t^k) =& \gamma_0 + \gamma_1(p_{t-1}^k - p_{t-1}^k) + \gamma_2 \Delta\pi_t^{jk} + \gamma_3 \Delta Z_t^{jk} \\ & + \mathbf{1}\{|p_{t-1}^j - p_{t-1}^k| \geq c\}(\delta_0 + \delta_1(p_{t-1}^j - p_{t-1}^k) + \delta_2 \Delta\pi_t^{jk} + \delta_3 \Delta Z_t^{jk}) + \varepsilon_t, \end{aligned} \tag{2.2.4}$$

where γ_0 and δ_0 are time trend coefficients, which may also be conceptually relevant. The parameters γ_0, γ_1, δ_0, and δ_1 reflect the degree of market integration. In particular, γ_1 and δ_1 represent the degree of 'error correction' characterizing departures from price parity, which

[1] If we define π_t^{jk} as the exchange rate of 1 unit of currency in market j to the currency in market k, a value of $\gamma_2 = -1$ represents perfect exchange rate pass-through.

are reflected in large values of $p^j_{t-1} - p^k_{t-1}$. The threshold parameter c represents the amount of proportional transaction costs that a price differential must exceed to cross the threshold and trigger the "trade" regime adjustments. We allow δ_0, δ_1,δ_2 and δ_3 to nonzero according to whether $|p^j_{t-1} - p^k_{t-1}|$ is within (i.e., $|p^j_{t-1} - p^k_{t-1}| < c$) or outside (i.e., $|p^j_{t-1} - p^k_{t-1}| \geq c$) of a symmetric band

Differencing is employed in this study to measure short-run relationships between variables. The first-difference model is utilized to avoid nonstationary variables, allowing a focus on immediate changes between variables. Differencing captures short-run dynamics, while the error correction process reflects longer-run relationships.

To assess the potential presence of transaction costs and other factors affecting price relationships, we consider a multivariate threshold distributed lag model that includes the price differential, exchange rate, and exogenous shocks as well as their lagged (past period) values, as follows:

$$\begin{aligned}\Delta(p^j_t - p^k_t) =& \gamma_0 + \gamma_1(p^j_{t-1} - p^k_{t-1}) + \sum_{l=0}^{L} \gamma_{2l} \Delta\pi^{jk}_{t-l} + \sum_{l=0}^{L} \gamma_{3l} \Delta z^{jk}_{t-l} \\ & + \mathbf{1}\{Q_t \geq c\} \left[\delta_0 + \delta_1(p^j_{t-1} - p^k_{t-1}) + \sum_{l=0}^{L} \delta_{2l} \Delta\pi^{jk}_{t-l} + \sum_{l=0}^{L} \delta_{3l} \Delta z^{jk}_{t-l} \right] + \varepsilon_t \\ & t = \{1, \ldots, T\},\end{aligned} \tag{2.2.5}$$

where L represents the maximum possible lag, which may increase with the sample size, slowly growing to infinity; Q_t denotes the lagged price differential used as the threshold variable, also known as the forcing variable for identifying thresholds, i.e., $Q_t \in \{p^j_{t-1} - p^k_{t-1}, \cdots, p^j_{t-L} - p^k_{t-L}\}$. We assume that the maximal lag order L is known. A distributed lag model (Almon (1965)) is utilized to reveal both short- and long-run dynamic effects between explanatory variables and response variables. Additionally, we employ LASSO, a flexible and supervised learning method. When dealing with time-lagged relationships, selecting the appropriate lag length is crucial in time series modeling. Typically, a well-defined lag length is chosen, and all lags up to that period are included in the model. However, in contexts like ours, where we investigate the dynamic relationship between price linkages, exchange rates, and market factors in agricultural commodities, the delivery time from one market to another spans several weeks to months. Consequently, not all lags are considered equally important in capturing price linkages in response to market shocks. In such scenarios, a distributed lag model (DLM) with lag selection, facilitated by LASSO, proves to be more suitable. LASSO's ability to determine distributed lags through a data-driven search enables a more precise representation of dynamic relationships in agricultural commodity markets. This framework offers a richer evaluation of price dynamics and patterns of adjustment.

Economic agents adjust their expectations of price differentials based on the level of transaction costs that pertain to previous periods. If the price differential exceeds certain thresholds, agents anticipate profitable gains from arbitrage and trade. The specified model offers the advantage of capturing simultaneous relationships between exchange rates and other relevant variables. Linear modeling techniques may not accurately capture the nonlinearities present in the model. Therefore, it is essential to investigate the impact of transaction costs on the market's response to an exchange rate shock or other market shocks nonlinearly. The existence of different levels of transaction costs can influence how price differentials respond to exchange rates or other shocks, as it determines the presence or absence of arbitrage opportunities. Indeed, a limitation of most existing threshold models of spatial price linkages lies in the typical assumption that transactions costs are constant (in levels or proportional terms). We allow transactions costs, which are inherently unobservable, to vary according to many conceptually relevant economic variables. The proposed model recognizes that the movements in the exchange rate can adjust how markets respond to changes, leading to different regimes based on transaction costs. By considering the effects of transaction costs, we can gain a more comprehensive understanding of the dynamics of the exchange rate pass-through mechanism and the effect of market factors.

The lag coefficients γ_s for $s = 1, \cdots L$ represent the lag distribution and define the pattern of how $\Delta\pi_{t-s}$ or Δz_{t-s} affects $\Delta(p_t^1 - p_t^2)$over time. The dynamic marginal effect of $\Delta\pi_t$ at the s-th lag is $\frac{\partial\Delta(p_t^j - p_t^k)}{\partial\Delta\pi_{t-s}} = \gamma_{1s}$. The dynamic marginal effect of $\Delta\pi_{t-s}^{jk}$ on $\Delta(p_t^j - p_t^k)$ at the s-th lag is given by $\frac{\partial\Delta(p_t^j - p_t^k)}{\partial\Delta\pi_{t-s}^{jk}} = \gamma_{1s}$. The dynamic marginal effect is essentially an effect of a temporary change in $\Delta\pi_{t-s}^{jk}$ on $\Delta(p_t^j - p_t^k)$, whereas the long-run cumulative effect $\sum_{s=1}^{L}\gamma_{1s}$ measures how much $\Delta(p_t^j - p_t^k)$ will be changed in response to a permanent change in $\Delta\pi$ when both$\Delta\pi_t$ and $\Delta(p_t^j - p_t^k)$ are stationary. The same derivation can be applied to any element of the vector Δz_{t-s}^{jk}. In the context of the threshold regression model considered here, γ_{1s} and γ_{2s} represent the effect regardless of the status of the forcing variable Q_t, termed the structural effect. On the other hand, δ_{1s}^{jk} and δ_{2s}^{jk} represent the effect when $Q_t > c$, referred to as the threshold effect.

Although a range of economic variables may be conceptually relevant to price linkages, the exact choice of variables and the resulting model specification is unclear. Transactions costs, local policies, and other economic phenomena may affect price linkages between international markets as well as between import and internal markets. Thus, we utilize data-driven methods to select an optimal specification from among a set of potentially relevant variables.

To obtain a specification that incorporates a broad range of variables in (2.2.5), we utilize a novel approach to inference and model selection: the debiasedLASSO (least absolute shrinkage and selection operator) method for high-dimensional threshold regression, which was recently developed by Yan (2023). This method allows us to fit the threshold regression models using the threshold LASSO estimator of Lee et al. (2016) in conjunction with the work of van de Geer et al. (2014). Compared to other estimators, this approach can construct asymptotically valid confidence bands for a low-dimensional subset of a high-dimensional parameter vector. Understanding the significance of the estimators can provide insights into the changes in transaction costs and threshold effects over time. However, standard approaches to inference are not applicable to such models.

To simplify, let

$$\alpha = (\gamma_0, \gamma_1, \gamma_{20} \cdots, \gamma_{2L}, \gamma_{30} \cdots, \gamma_{3L}, \delta_0, \delta_1, \delta_{20} \cdots, \delta_{2L}, \delta_{30} \cdots, \delta_{3L})'$$

be slope parameter vector, The dimension of α is $4+2(1+p)(L+1)$, where p is number of other exogenous shocks. Let $\boldsymbol{X}$ be a $T \times [2+(1+p)(L+1)]$ matrix of all regressors. To provide a more precise description of our estimation procedures, we propose a three-step estimation approach for the model. The three-step procedure can be outlined as follows:

Step 1.

For each $c \in \mathbb{C}$, $\widehat{\alpha}(c)$ is defined as

$$\widehat{\alpha}(c) := \operatorname{argmin}_{\alpha} \left\{ T^{-1} \sum_{t=1}^{T} \left(\Delta(p_t^j - p_t^k) - [X_t', X_t' \mathbf{1}\{Q_t \geq c\}]' \alpha \right)^2 + \lambda \| \boldsymbol{D}(c) \alpha \|_1 \right\}, \tag{2.2.6}$$

where we can rewrite the ℓ_1 penalty as

$$\lambda |\boldsymbol{D}(c)\alpha| 1 = \lambda \sum j = 1^{2+(1+p)(L+1)} \left[\left\| X^{(j)} \right\|_n \left| \alpha^{(j)} \right| + \left\| X^{(j)}(\tau) \right\|_n \left| \alpha^{(1+(1+p)(L+1)+j)} \right| \right],$$

to adjust the penalty differently for each coefficient, depending on the scale normalizing factor. The tuning parameter λ can be selected using either the Akaike Information Criterion (AIC) or the Schwarz Information Criterion (SBC).

Define $\widehat{c}$ as the estimate of c_0 such that:

$$\widehat{c} := \operatorname{argmin}_{c \in \mathbb{C} \subset \mathbb{R}} \left\{ T^{-1} \sum_{t=1}^{T} \left(\Delta(p_t^1 - p_t^2) - [X_t', X_t' \mathbf{1}\{Q_t \geq c\}]' \widehat{\alpha}(c) \right)^2 + \lambda \| \widehat{\alpha}(c) \|_1 \right\}. \tag{2.2.7}$$

Following Yan (2023), we next turn to variable selection utilizing thresholding. We follow sharp threshold detection techniques provided by Callot et al. (2017) to finding out whether there is a threshold or not, that is, whether

$$(\delta_0, \delta_1, \delta_{20} \cdots, \delta_{2L}, \delta_{30} \cdots, \delta_{3L})'$$

is nonzero or not.

Step 2. We define the thresholded LASSO estimator as

$$\tilde{\delta}_{(j)}(\hat{c}) = \begin{cases} \hat{\delta}_{(j)}(\hat{c}), & \text{if } |\hat{\delta}_{(j)}(\hat{c})| \geq H, \\ 0, & \text{if } |\hat{\delta}_{(j)}(\hat{c})| < H. \end{cases} \tag{2.2.8}$$

where H is the threshold determining whether a coefficient should be classified as zero or nonzero and $\hat{\delta}^{(j)}(\hat{c})$ are elements of the LASSO estimator defined by (2.2.6) and (2.2.7) jointly. In particular, we shall see that choosing $H = 2D\lambda$ yields consistent model selection. The thresholding parameter D can be selected using AIC or SBC through grid search. This ensures that parameters smaller (in absolute value) than $\widehat{D}\hat{\lambda}$ are set to zero by the thresholded LASSO.

The thresholded LASSO in (2.2.8) can achieve threshold selection consistency. The consistency of the LASSO estimator implies that if the underlying true model is nonlinear, then the LASSO estimator will correctly estimate any of the non-zero parameters, including $(\delta_0, \delta_1, \delta_{20}, \cdots, \delta_{2L}, \delta_{30}, \cdots, \delta_{3L})$. In other words, if any of these parameters are non-zero, the LASSO estimator will consistently estimate them as non-zero, indicating the presence of a nonlinear relationship between the variables. This is in contrast to the conventional 'self-exciting' threshold autoregressive (SETAR) model, where nonlinear tests such as Hansen's modification of standard Chow-type tests, Tsay (1989) linearity test, or neural network tests of linearity are utilized to detect nonlinearity. Therefore, if we misspecify a linear model and use the LASSO method for the threshold model described here, we may estimate all threshold effects as zero for a sufficiently large sample size. To put it another way, if our estimates of $(\delta_0, \delta_1, \delta_{20}, \cdots, \delta_{2L}, \delta_{30}, \cdots, \delta_{3L})$ after steps 1 and 2 have at least one non-zero, it indicates that the probability of the model being linear approaches 0.

Once variables are selected through LASSO estimation and the presence of threshold effects is confirmed, the shrinkage bias induced by penalization in the LASSO loss function becomes evident in the properly scaled limiting distribution of the LASSO estimator. Therefore, to enable statistical inference, an estimation strategy must be employed to eliminate this bias. However, when modeling threshold regression with a rich set of variables, a challenge emerges. Threshold models entail splitting the sample based on a continuously-distributed variable. With a rich set of regressors, there's a risk that the number of observations in any split sample may be less than the number of variables, leading to a reduced-rank sample covariance matrix. Standard approaches are inadequate in such a situation. To desparsify (debias) our LASSO estimator, an approximate inverse of a certain singular sample covariance matrix is needed, as discussed by van de Geer et al. (2014). For a more in-depth exploration and extensions in the case of the LASSO applied to the high-dimensional threshold regression model, detailed information can be found in Yan (2023). However, we do not delve further into these extensions here.

Step 3

Finally, we can obtain debiasedLASSO estimates for the threshold model, which is given by:

$$\hat{a}(\widehat{c}) = \hat{\alpha}(\hat{c}) + \widehat{\mathbf{\Theta}}(\widehat{c})\boldsymbol{X}'(\widehat{c})(\Delta(p^1 - p^2) - \boldsymbol{X}(\widehat{c})\widehat{\alpha}(\widehat{c}))/n, \tag{2.2.9}$$

where

$$\widehat{\mathbf{\Theta}}(\widehat{c}) = \begin{bmatrix} \widehat{\boldsymbol{B}}(\widehat{c}) & -\widehat{\boldsymbol{B}}(\widehat{c}) \\ -\widehat{\boldsymbol{B}}(\widehat{c}) & \widehat{\boldsymbol{A}}(\widehat{c}) + \widehat{\boldsymbol{B}}(\widehat{c}) \end{bmatrix}, \tag{2.2.10}$$

and $\widehat{\boldsymbol{B}}(\widehat{c})$ and $\widehat{\boldsymbol{A}}(\widehat{c})$ are the inverse or approximate (if the sample covariance matrix is singular) inverse of the split sample covariance matrices.

For model selection (i.e. to determine the optimal lag structure on forcing variable Q_t), one can use selection criteria such as AIC or SBC to select the optimal lag structure for the forcing variables.

2.3 Empirical Application

The empirical analyses in our study focus on international corn markets, specifically three major exporting markets: the US, Argentina, and Ukraine. Despite its widespread consumption and spatial dispersion, corn production is typically concentrated in specific regions. These three markets collectively accounted for 66.5% of the world corn trade by volume in the 2021/2022 marketing year. Given the intricate spatial dynamics of the corn market, analyzing spatial linkages is crucial for an understanding of the underlying market dynamics and overall performance and behavior.

We collected monthly maize prices and other relevant data from multiple sources which are discussed below. As noted above, the main dependent variable of interest in this study is the maize price in international markets. We collected the yellow corn export prices of the US, Ukraine, and Argentina. Price data for the main three export markets were obtained from the FAO Food Price Monitoring and Analysis (FPMA) Tool, reporting prices in US dollars per metric ton.

Our dataset spans from April 2002 to December 2022, providing 243 monthly observations for each series. However, due to data availability constraints, market factors data for Ukraine is only accessible from April 2002 to February 2022, comprising 239 observations. Similarly, market factors data for Argentina is available from July 2003 to December 2022, encompassing 234 observations. To address missing values, we applied spline interpolation during the relevant periods where prices were missing.[2]

We collected exchange rates for Ukraine (United States dollar(USD) to Ukrainian Hryvnia(UAH)) and Argentina (USD to Argentine peso(ARS)). Additionally, we collected the Baltic Exchange Dry Index, measuring the cost of shipping dry goods like maize worldwide. To capture US market factors, we sourced data from the Federal Reserve Economic Data (FRED), including unemployment rates, the consumer price index, the industrial production index, interest rates, and gasoline prices. For US corn stocks data, we utilized quarterly information from the US Feed Grains Yearbook, converting it into monthly data for analysis[3] Market factors for Ukraine, such as unemployment rates, the consumer price index, and the industrial production index, along with those for Argentina, such as unemployment rates, consumer price index, and inflation rate, were sourced from the National Summary Data Pages (NSDPs)[4]

The basic unit of analysis used throughout is the natural logarithm of the price ratio, denoted as $p_t^j - p_t^k (= \ln(P_t^j / P_t^k))$, where i and j indicate locations (i.e., $j, k = 1, 2, 3$ denote the US, Ukraine, and Argentina respectively), and t is a time index such that $t = 1, \cdots, T$. The international price data are shown in logarithmic form in Figure 2.1b.

Figure 2.2 presents a graphical representation of logarithmic pairs of prices plotted against each other, offering insights into the relationship between price levels and price differentials. Deviations from the 45-degree line in each plot reveal distinct basis patterns, where one price tends to be higher or lower than the other. These patterns likely reveal the influence of transaction costs associated with regionally distinct market trades. While these countries are exporters only, in the market integration framework, maize flows between the three markets

[2]Cubic spline interpolation was employed to proxy missing price data within continuous periods. 20 observations are missing from April 2002 to February 2022 for the Ukrainian Maize export price.

[3]To align the data frequencies for our econometric analysis, cubic spline interpolation was applied to convert the quarterly US corn beginning stock data into the same frequency as all other monthly variables. US corn beginning stock data is from 2001 Q2 (Dec-Feb) to 2021 Q3 (Mar-May), totaling 82 observations and converted to 246 monthly data. The seasonality of corn stocks was respected in our interpolation methodology.

[4]To align the data frequencies for our econometric analysis, cubic spline interpolation was applied to convert the quarterly Argentina unemployment rate and Ukraine employment rate into the same frequency as all other monthly variables. The data for Ukraine unemployment ranges from 2022 Q1 to 2021 Q4 (standard calendar quarters), totaling 80 observations and converted to 240-month observations. The data for Argentinian unemployment spans from 2002 Q4 to 2022 Q4, comprising 81 observations and converted to 243-month observations. Given that the variables, including the consumer price index of Ukraine, industrial production index of Ukraine, and industrial production index of Argentina, are segmented into multiple partitions over the selected period, and each partition is calculated using different units in the data sources, we employ cubic spline interpolation to estimate the data for the months where unit changes occur. Note that these variables are typically not volatile on a month-to-month basis, making spline interpolation a reasonable approach to converting the data to a monthly basis.

can occur in any direction, depending on potentially profitable arbitrage opportunities. Our observations from the figures indicate that situations where the price of Ukrainian maize surpasses the prices of US maize and Argentina maize occur more frequently.

To examine the characteristics of time series prices and identify the most appropriate model for evaluating spatial price linkages, we conducted augmented Dickey-Fuller tests for each pair of price differentials. The results of the Augmented Dickey-Fuller (ADF) tests for the stationarity of the price differentials are presented in Table B.1 in the appendix, which indicates that the null hypothesis of nonstationarity of the price differentials is strongly rejected in every case. This is as expected since a nonstationary differential would imply that prices can drift arbitrarily far apart.

Transmission elasticities($\frac{\partial P_t^j}{\partial P_t^k}$) close to one provide support for market integration, with 1.0 corresponding to perfect market integration. Additionally, we performed ADF tests on the first differences of the logarithms of variables (all variables are logarithmic except for the unemployment rates of three countries and the inflation rate of Argentina), exchange rates, and other exogenous shocks. The results, presented in Table B.2 in the appendix, indicate that all these variables significantly differ from nonstationary series. Our Augmented Dickey-Fuller (ADF) test on the first differences of all variables strongly rejects the null hypothesis of nonstationarity. Therefore, we can confidently implement Equation (2.2.5) for estimating the model with the available data.

Before consideration of two-regime switching models, we consider a suite of tests intended to detect departures from linearity in conventional time-series models. A range of (non-) linearity tests were conducted for the price data. We applied a standard Self-Exciting Threshold AutoRegressive (SETAR) model, as formulated by Goodwin and Piggott (2001), to prices in spatially distinct markets for each of the market pairs. The specification is given by:

$$\Delta(p_t^j - p_t^k) = \gamma_0 + \gamma_1(p_{t-1}^j - p_{t-1}^k) + \mathbf{1}\{Q_t \geq c\}\left[\delta_0 + \delta_1(p_{t-1}^j - p_{t-1}^k)\right] + \varepsilon_t \tag{2.3.1}$$

where c is a threshold parameter, and $\gamma_1 + \delta_1$ is the parameter for trade regime. Each of the linearity tests was applied to the collection of prices. Tests on pairs of prices were conducted on the differential between logarithmic prices. Nonlinearity testing results are contained in Table 2.1.[5] The tests for all international market pairs reject linearity in at least one of the alternative linearity tests at a 10% significance level. Thus, these tests robustly reject linearity among the price linkages, prompting the exploration of alternative, flexible specifications capable of accommodating nonlinearities. Threshold models are a likely candidate for a nonlinear representation of the price relationships.

[5]Hansen's modification of standard Chow-type tests of the bootstrapping results presented in this paper utilized 1000 replications.

The question remains as to the most appropriate specification of the alternative models of price parity. We have suggested that, despite the fact that prices are all quoted in US dollar terms, exchange rates may nevertheless play a role in international price linkages. Specifically, if import prices pertain to an intermediary step in trade between internal markets, where different currencies may exist, and international markets, exchange rates may still be relevant to the price linkages. If exchange rates are found to exert a statistically significant effect on price linkages, exchange rate over- or under-shooting may exist. Further, it is unclear as to whether additional variables may be relevant to price linkages. Markets are separated by unobservable transactions costs, which may in turn be influenced by other economic variables. Hence, we utilize LASSO methods to select an optimal specification.

We initially estimate price relationships for the three pairs of market prices using a standard autoregressive model of the form:

$$\Delta(p_t^j - p_t^k) = \gamma_0 + \gamma_1(p_{t-1}^j - p_{t-1}^k), \tag{2.3.2}$$

where $(j,k) = \{(1,2),(1,3),(2,3)\}$, with the indices representing the US, Ukraine, and Argentina as $1,2,3$ respectively. γ_0 and γ_1 are parameters reflecting the degree of market integration. In particular, we expect a small but negative value of γ_1, so the price differential $p_t^j - p_t^k$ converges to 0 at the rate of $1+\gamma_1 < 1$. A value of γ_1 closer to zero implies a slower adjustment to shocks.

Moreover, we then extend our analysis to include the exchange rate, considering the following specification:

$$\Delta(p_t^1 - p_t^2) = \gamma_0 + \gamma_1(p_{t-1}^1 - p_{t-1}^2) + \gamma_2 \Delta\pi_t^{12}, \tag{2.3.3}$$

Besides model (2.3.2) and (2.3.3), estimations are conducted based on model (2.3.1). In addition, we use exchange rates as covariates and estimate threshold models of the form:

$$\begin{aligned}\Delta(p_t^j - p_t^k) =& \gamma_0 + \gamma_1(p_{t-1}^j - p_{t-1}^k) + \gamma_2 \Delta\pi_t^{jk} \\ &+ \mathbf{1}\{Q_t \geq c\}\left[\delta_0 + \delta_1(p_{t-1}^j - p_{t-1}^k) + \delta_2 \Delta\pi_t^{jk}\right] + \varepsilon_t.\end{aligned} \tag{2.3.4}$$

The 2nd to 5th columns in Table 2.2 present estimates of two standard autoregressive price parity models, as denoted by equations (2.3.2) and (2.3.3), and estimates of two threshold autoregressive price parity models, referenced by equations (2.3.1) and (2.3.4). The 2nd to 5th columns in Table 2.3 and Table 2.4 denote the same models for the other two pairs of market prices. When threshold behavior is disregarded in the linear model, in every case, models incorporating exchange rates suggest adjustments in response to deviations from equilibrium that are at least as fast as the models when exchange rates are ignored. However, when considering the threshold models, the estimations for the "no-trade" regime are positive but close to zero in every case except for US/Ukraine with the exchange rate. The theoretical expectation is for these estimates to be negatively approaching zero. Nevertheless, except for the model of Ukraine/Argentina with the exchange rate, we obtain negative estimates for δ_1, indicating much faster adjustments in response to deviations from equilibrium in the "trade" regime than the "no-trade" regime. This is implied if the estimates of structural effects γ_1 are negative. It's interesting to note that when we consider the exchange rate effect in the models, the exchange rates for models of US/Ukraine and Ukraine/Argentina, with or without a threshold, exhibit undershooting. Specifically, the exchange rate effect is a perfect pass-through in the model of US/Argentina without a threshold, while there is no significant effect in the model of US/Argentina with a threshold.

As mentioned earlier, our estimation procedures for threshold regression offer the advantage of variable selection and threshold detection, eliminating the need for conventional nonlinear tests commonly used in threshold models. In our study, the covariates included in all market pairs are the exchange rate, the Baltic Exchange Dry Index, unemployment rates, and industrial production indexes for each market. Additionally, for pairs including the US, we utilize US interest rates, US Corn Stock, and US gas prices as control variables. Due to data availability, consumer price indexes are collected for the US and Ukraine, while inflation rates are used for Argentina.

To determine the optimal lag structure for the forcing variable, we select the lag order with the lowest SBC value for each model. The candidates for lag orders range from 1 to 6, considering the price differentials for pairs of market prices. Table B.3 presents the SBC values for the threshold LASSO estimation, which is used to select the lag structure for the forcing variable Q_t. In this context, the symmetric lagged price differential $|p^j_{t-d} - p^k_{t-d}|$ transforms into Q_t, representing the quantile of $|p^j_{t-d} - p^k_{t-d}|$ in selected samples. The estimation of thresholds is conducted using a grid search. An assumption is made that all $|p^j_{t-d} - p^k_{t-d}|$ values are distinct. This is a convenient condition, ensuring that the transformation into quantiles is a one-to-one function without any loss of generality. This assumption holds under the assumption of continuous distribution for $|p^j_{t-d} - p^k_{t-d}|$. As the AIC tends to produce less sparse solutions overall, and SBC applies a stronger penalty on the degrees of freedom, it is more conservative in variable selection compared to AIC. Therefore, we use SBC to select the optimal lag structure and AIC for tuning the parameter λ.

As shown in Table B.3, the forcing variables for pairs of market prices of the US/Ukraine and US/Argentina are selected as the 4-month lagged price differentials, while a 3-month lag is chosen for Ukraine/Argentina. The threshold estimates offer insights into transaction costs. Simultaneously, the quantile estimates (refer to the same tables) illuminate whether, during the selected periods, monthly observations more frequently align with trade regimes characterized by lower quantile estimates. Based on the optimal lag structure determined by minimum BIC values, in the scenarios of US/Ukraine and Ukraine/Argentina, price differentials within the bands occur more frequently, as quantile estimates of the threshold parameters exceed 0.5. However, for US/Argentina, arbitrage activities are triggered more frequently, leading to the "trade" regime. When examining the magnitude of the price differential estimates, the width of the band representing "no trade", as implied by the thresholds, is widest for the Ukraine/Argentina markets and narrowest for the US/Argentina markets.

The last (6th) column in each of the tables (Table 2.2, Table 2.3, and Table 2.4) presents estimates of the threshold model using debiased LASSO, as based on equation (2.2.5). The optimal lag structures for the forcing variables are the same results as presented in Table B.3. In the case of US/Ukraine and US/Argentina, the estimates of adjustment in response to deviations from equilibrium (γ_1 and δ_1) are negative and close to zero, showing a significantly faster rate in the "trade" regime. However, in the case of US/Argentina and Ukraine/Argentina, there is no significant difference between the estimates of adjustment in response to deviations from equilibrium in the two regimes.In all cases regarding the exchange rate pass-through effect, there are no significant results indicating overshoot or undershoot, and there is no significant difference between the two regimes.

Tables B.4 display the signs of statistically significant estimates of market factors at the 10% significance level, except for the error correction rate estimates (γ_0, γ_1, δ_0, and δ_1) and exchange rate effect estimates (γ_2 and δ_2), corresponding to models represented by the last column in each of the tables (Table 2.2, Table 2.3, and Table 2.4). It's worth noting that the debiased LASSO estimates are considered statistically insignificant when the slope estimates by Equations (2.2.6) and (2.2.7) are zero. Therefore, only non-zero estimates by Equations (2.2.6) and (2.2.7) are presented.

The standard threshold model assumes a fixed threshold, a potentially limiting assumption. It is reasonable to consider that relationships may evolve, signaling structural changes in the underlying economic dynamics. To explore this possibility, we introduce partitions that reflect changes in market environments. The data is segmented into two periods corresponding to two significant economic shocks: the 2014 Crimean crisis (e.g., Korovkin and Makarin (2023)) and the global financial/economic crisis of 2008-09 (e.g., Liefert et al. (2021)). Specifically, the breakpoints for these events are defined as February 2014 and October 2008, respectively, in our monthly dataset. The corresponding optimal lag structures for the forcing variable of threshold LASSO estimation are provided in Table B.5, B.6, and B.7. Introducing a break in the dataset corresponding to October 2008 for the global financial/economic crisis and February 2014 for the 2014 Crimean crisis reveals that, in most cases, the selected optimal lagged forcing variables differ across the entire period, pre-breakpoint, and post-breakpoint. Profitable arbitrage opportunities are more frequent only in the post-February 2014 period for the US/Ukraine markets, while in every other case, such opportunities are fewer. When examining the magnitude of the price differential estimates, post-break threshold bands are narrower in all comparisons except for US/Argentina pre/post-October 2008. Subsequently, we remove the shrinkage bias introduced by the penalization in Equation (2.2.6) using Equation (2.2.9) for post-selection statistical inference. Our estimation setup considers a richer examination of price linkage among global maize markets. The fundamental framework of the threshold model illustrates that if any of the estimates of the slope coefficients (exchange rate pass-through or exogenous shock) are regime-specific, the effects of certain lagged exchange rate or exogenous shock on price differentials (which could be lagged variables) between two distinct markets differs depending on the magnitude of a certain forcing variable representing unobserved transaction costs. Estimates of non-zero differences between the two regimes imply nonlinear relationships. The slope coefficient directly corresponds to elasticity, measuring the responsiveness of the dependent variable (the price linkages in time t) to changes in the explanatory factors (lagged exchange rate between the two markets or any market factor). A straightforward way to illustrate the effects of exchange rates, market factors, or exogenous shocks on potential deviations from price parity is by analyzing the coefficient estimates obtained from our estimations. All lagged variables are allowed to have a dynamic linear effect or a dynamic nonlinear effect depending on the existence of a regime switch (threshold).

Tables B.8, B.9, and B.10 offer a comprehensive summary of the estimates for the degree of error correction and exchange rate effects based on (2.2.5) using the LASSO method[6] These tables offer valuable insights into the adjustments and effects in each market pair, illuminating the interdependence between different markets. In almost every case, spanning entire periods and structural breaks, the estimates of the degree of “error correction" adjustments approach negative zero, except for the Post-October 2008 period in the US/Ukraine markets. The threshold models suggest adjustments in response to deviations from equilibrium in the “trade" regime that are at least as fast as those in the “no-trade" regime, except for the Post-October 2008 and Post-February 2014 periods in the US/Ukraine markets, and the Post-October 2008 period in the US/Argentina market, as well as the Pre-February 2014 period in the Ukraine/Argentina market. The threshold models indicate that exchange rates exhibit perfect pass-through to markets during the Post-October 2008, Pre-February 2014, and throughout the entire period in the US/Argentina pair across all 11 models.

[6]If γ_1, γ_2, δ_1, or δ_2 is not selected by the LASSO step, their estimates and standard errors are left blank in the table. The detailed debiased LASSO estimates for all variables are not included here due to space constraints but are presented in the Appendix.

In every case, some other market factors and their past period values are selected and tested as statistically significant.

2.4 Summary and Concluding Remarks

We develop a model of price parity in spatially distinct international export markets for maize to investigate the degree of “error correction”, the exchange rate pass-through, and other market factor effects. The models are developed within the framework of high-dimensional threshold models. We consider such nonlinear models, that has developed an increasingly rich set of factors in models of spatial market integration, as extensions to existing literature. The debiased LASSO estimation procedures are used to specify the models.

In summary, our findings consistently indicate faster adjustments in response to deviations from equilibrium in conditions of profitable trade and arbitrage compared to the case of no trade. The markets exhibit strong linkages in most cases, with confirmed nonlinear adjustments. Aligned with existing research, there is insufficient evidence to reject perfect pass-through of exchange rate fluctuations. However, the process of price convergence to market equilibrium is influenced by changes in some market factors, such as transportation cost, market unemployment rate, inflation, and the index of industrial production. These differences signify more substantial disequilibrium conditions, thereby presenting larger arbitrage opportunities.

Nonlinearity test				
	US/ Ukraine		US/Argentina	
	Test Statistics	p-value	Test Statistics	p-value
Teraesvirta's neural network test χ^2	2.596	0.273	2.541	0.281
White neural network test χ^2	2.898	0.235	6.587	0.037
Keenan's one-degree test for nonlinearity F-test	1.639	0.202	0.115	0.735
Tsay's Test for nonlinearity F-test	1.360	0.099	1.125	0.340
Likelihood ratio test for threshold nonlinearity χ^2	22.777	0.077	10.476	0.132
(SETAR) models: Linear AR versus 1 threshold TAR F-test	5.882	0.515	8.795	0.202
	Ukraine/Argentina			
	Test Statistics	p-value		
Teraesvirta's neural network test χ^2	3.005	0.223		
White neural network test χ^2	3.529	0.171		
Keenan's one-degree test for nonlinearity F-test	2.181	0.141		
Tsay's Test for nonlinearity F-test	2.166	0.048		
Likelihood ratio test for threshold nonlinearity χ^2	10.573	0.207		
(SETAR) models: Linear AR versus 1 threshold TAR F-test	6.216	0.479		

Table 2.1: Nonlinearity Specification Testing Results

Parameter	Linear without exchange rate	Linear with exchange rate	Threshold without exchange rate	Threshold with exchange rate	Threshold using debiased LASSO
(Intercept) γ_0	-0.008*	-0.010**	-0.006	-0.006	-0.018 **
	(0.005)	(0.005)	(0.005)	(0.005)	(0.008)
degree of "error correction" γ_1	-0.137 ***	-0.142 ***	0.040	-0.091	-0.090 ***
	(0.032)	(0.032)	(0.071)	(0.086)	(0.023)
exchange rate effect γ_2		0.234**		0.083	-0.023
		(0.100)		(0.112)	(0.094)
(Intercept) δ_0			-0.009	-0.021 **	
			(0.011)	(0.010)	
degree of "error correction" δ_1			-0.230**	-0.108	-0.070 *
			(0.082)	(0.094)	(0.041)
exchange rate effect δ_2				0.787***	0.300
				(0.242)	(0.223)
threshold estimate			0.130	0.111	0.089
threshold quantile			0.73	0.65	0.55
threshold time delay			1	1	4
Observations	233	233	233	233	233
R^2	0.074	0.095	0.105	0.142	···
Adjusted R^2	0.070	0.087	0.089	0.120	···
F-statistic	18.44	12.12	6.71	6.28	···
AIC	-5.486	-5.501	-5.503	-5.528	-5.502

Note: *p<0.1; **p<0.05; ***p<0.01

Table 2.2: Model Estimates: US/ Ukraine

Parameter	Linear without exchange rate	Linear with exchange rate	Threshold without exchange rate	Threshold with exchange rate	Threshold using debiased LASSO
(Intercept) γ_0	0.001	0.001	0.003	0.002	0.009
	(0.002)	(0.003)	(0.004)	(0.005)	(0.008)
degree of "error correction" γ_1	-0.136 ***	-0.136 ***	0.278	0.274	
	(0.033)	(0.034)	(0.221)	(0.222)	
exchange rate effect γ_2		0.000		0.048	
		(0.071)		(0.130)	
(Intercept) δ_0			-0.001	-0.000	
			(0.005)	(0.006)	
degree of "error correction" δ_1			-0.424	-0.420*	-0.160***
			(0.223)	(0.225)	(0.023)
exchange rate effect δ_2				-0.076	
				(0.155)	
threshold estimate			0.033	0.033	0.043
threshold quantile			0.36	0.36	0.46
threshold time delay			1	1	4
Observations	228	228	228	228	228
R^2	0.068	0.068	0.083	0.052	···
Adjusted R^2	0.064	0.060	0.067	0.025	···
F-statistic	16.57	8.249	5.079	1.936	···
AIC	-6.593	-6.584	-6.591	-6.575	-6.650

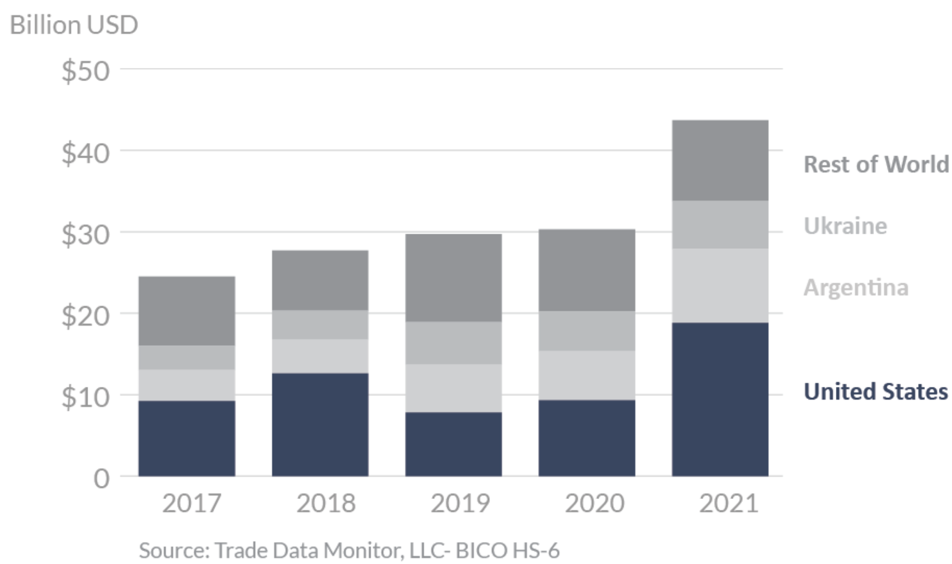
Note: *p<0.1; **p<0.05; ***p<0.01

Table 2.3: Model Estimates: US/ Argentina

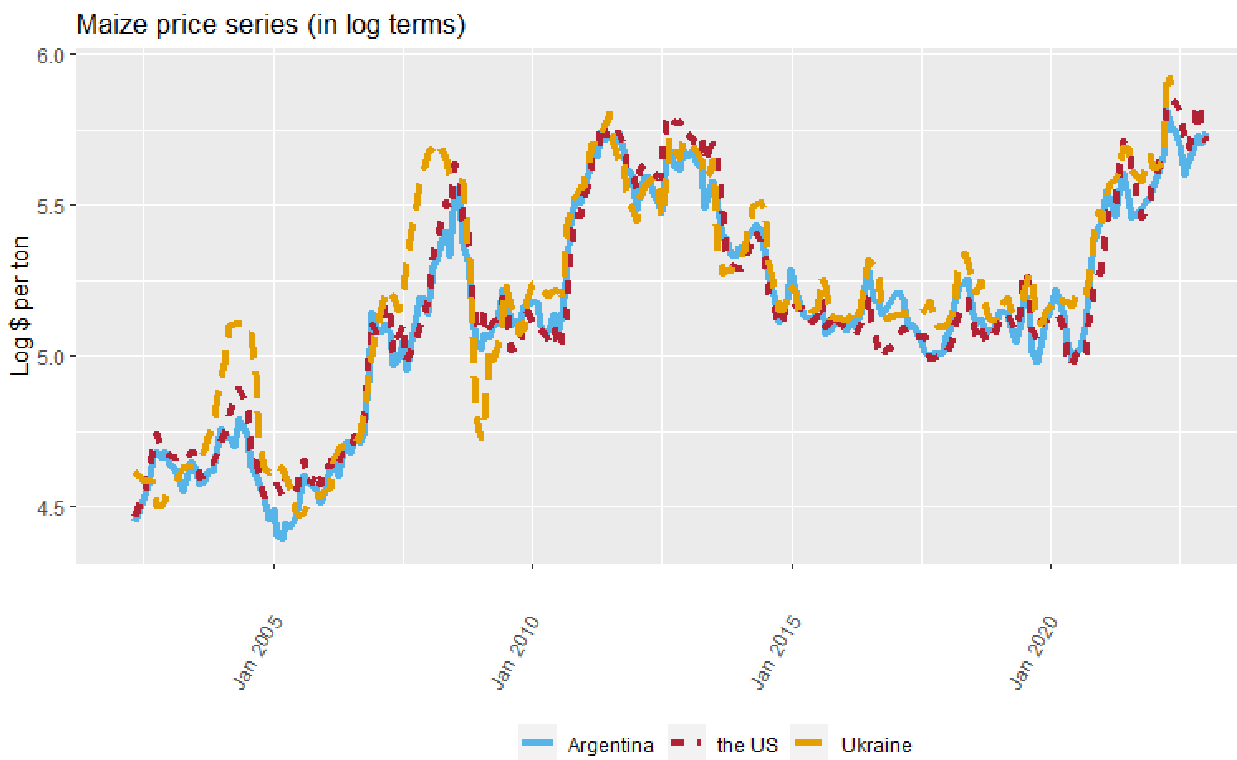
Parameter	Linear without exchange rate	Linear with exchange rate	Threshold without exchange rate	Threshold with exchange rate	Threshold using debiased LASSO
(Intercept) γ_0	-0.009*	-0.008	0.002	-0.003	0.087
	(0.005)	(0.005)	(0.008)	(0.005)	(0.093)
degree of "error correction" γ_1	-0.147***	-0.152***	0.131	0.093	-0.139***
	(0.035)	(0.034)	(0.275)	(0.076)	(0.020)
exchange rate effect γ_2		0.178**		0.197**	0.473
		(0.082)		(0.093)	(0.410)
(Intercept) δ_0			-0.019*	0.012	
			(0.010)	(0.011)	
degree of "error correction" δ_1			-0.305	0.052	
			(0.278))	(0.142)	
exchange rate effect δ_2				-0.242	9.051
				(0.220)	(8.443)
threshold estimate			0.046	0.113	0.134
threshold quantile			0.40	0.71	0.77
threshold time delay			1	1	3
Observations	218	218	218	218	218
R^2	0.074	0.094	0.089	0.052	···
Adjusted R^2	0.069	0.085	0.072	0.025	···
F-statistic	17.2	11.1	5.252	1.936	···
AIC	-5.469	-5.481	-5.467	-5.408	-5.658

Note: *p<0.1; **p<0.05; ***p<0.01

Table 2.4: Model Estimates: Ukraine/ Argentina



(a)



(b)

Figure 2.1: (a) Global Corn Exports by Country and Marketing Year, Source: U.S. Department of Agriculture, Foreign Agricultural Service (2022). (b) Maize Retail Price Series (in log terms) by Country

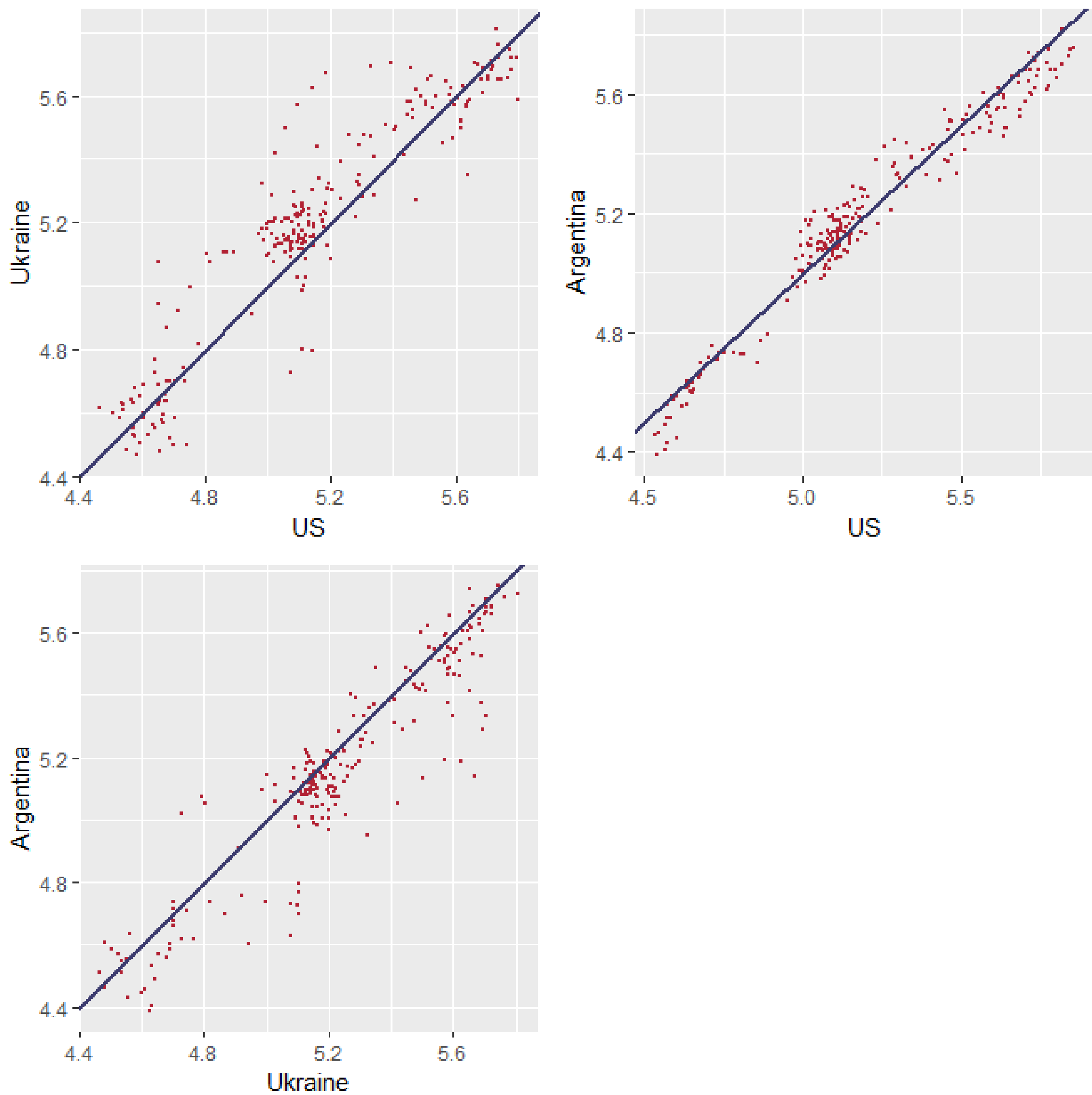


Figure 2.2: Maize Market Prices Pairs (in logarithms)

REFERENCES

Adamek, R., Smeekes, S., and Wilms, I. (2023). Lasso inference for high-dimensional time series. *Journal of Econometrics*, 235(2):1114–1143.

Afonso, A. and Jalles, J. (2013). Growth and productivity: The role of government debt. *International Review of Economics & Finance*, 25(C):384–407.

Almon, S. (1965). The distributed lag between capital appropriations and expenditures. *Econometrica*, 33(1):178–196.

Banerjee, A., Dolado, J. J., Galbraith, J. W., and Hendry, D. (1993). *Co-integration, Error Correction, and the Econometric Analysis of Non-Stationary Data*. Oxford University Press.

Barrett, C. B. and Li, J. R. (2002). Distinguishing between equilibrium and integration in spatial price analysis. *American Journal of Agricultural Economics*, 84(2):292–307.

Belloni, A., Chernozhukov, V., and Hansen, C. (2014). Inference on Treatment Effects after Selection among High-Dimensional Controls†. *The Review of Economic Studies*, 81(2):608–650.

Belloni, A., Chernozhukov, V., and Wei, Y. (2016). Post-selection inference for generalized linear models with many controls. *Journal of Business & Economic Statistics*, 34(4):606–619.

Belloni, A., Hansen, C., and Newey, W. (2019). Simultaneous confidence intervals for high-dimensional linear models with many endogenous variables.

Bickel, P. J., Ritov, Y., and Tsybakov, A. B. (2009). Simultaneous analysis of Lasso and Dantzig selector. *The Annals of Statistics*, 37(4):1705 – 1732.

Cai, T., Liu, W., and Luo, X. (2011). A constrained l_1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*, 106(494):594–607.

Callot, L., Caner, M., Kock, A. B., and Riquelme, J. A. (2017). Sharp threshold detection based on sup-norm error rates in high-dimensional models. *Journal of Business & Economic Statistics*, 35(2):250–264.

Caner, M. (2023). Generalized linear models with structured sparsity estimators. *Journal of Econometrics*, 236(2):105478.

Caner, M. and Hansen, B. E. (2004). Instrumental variable estimation of a threshold model. *Econometric Theory*, 20(5):813–843.

Caner, M. and Kock, A. B. (2018). Asymptotically honest confidence regions for high dimensional parameters by the desparsified conservative lasso. *Journal of Econometrics*, 203(1):143–168.

Caner, M. and Kock, A. B. (2019). High dimensional linear gmm.

Caner, M., Medeiros, M., and Vasconcelos, G. F. (2023). Sharpe ratio analysis in high dimensions: Residual-based nodewise regression in factor models. *Journal of Econometrics*, 235(2):393–417.

Chambers, R. G. and Just, R. E. (1981). Effects of exchange rate changes on u.s. agriculture: A dynamic analysis. *American Journal of Agricultural Economics*, 63(1):32–46.

Chang, J., Qiu, Y., Yao, Q., and Zou, T. (2018). Confidence regions for entries of a large precision matrix. *Journal of Econometrics*, 206(1):57–82.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2017). Central limit theorems and bootstrap in high dimensions. *The Annals of Probability*, 45(4):2309–2352.

Chiang, H. D., Rodrigue, J., and Sasaki, Y. (2023). Post-selection inference in three-dimensional panel data. *Econometric Theory*, 39(3):623–658.

Chudik, A., Mohaddes, K., Pesaran, M. H., and Raissi, M. (2017). Is there a debt-threshold effect on output growth? *The Review of Economics and Statistics*, 99(1):135–150.

Dudley, R. M. (1985). An extended wichura theorem, definitions of donsker class, and weighted empirical distributions. In Beck, A., Dudley, R., Hahn, M., Kuelbs, J., and Marcus, M., editors, *Probability in Banach Spaces V*, pages 141–178, Berlin, Heidelberg. Springer Berlin Heidelberg.

Fan, J., Liao, Y., and Mincheva, M. (2011). High-dimensional covariance matrix estimation in approximate factor models. *The Annals of Statistics*, 39(6):3320 – 3356.

Gold, D., Lederer, J., and Tao, J. (2020). Inference for high-dimensional instrumental variables regression. *Journal of Econometrics*, 217(1):79–111.

Goodwin, B. K., Grennes, T., and Wohlgenant, M. K. (1990). Testing the law of one price when trade takes time. *Journal of International Money and Finance*, 9(1):21–40.

Goodwin, B. K., Holt, M. T., and Prestemon, J. P. (2021). Semi-parametric models of spatial market integration. *Empirical Economics*, 61:2335–2361.

Goodwin, B. K. and Piggott, N. E. (2001). Spatial market integration in the presence of threshold effects. *American Journal of Agricultural Economics*, 83(2):302–317.

Grennes, T., Caner, M., and Koehler-Geib, F. (2010). *Finding The Tipping Point – When Sovereign Debt Turns Bad*. The World Bank.

Hansen, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica*, 64(2):413–430.

Hansen, B. E. (2000). Sample splitting and threshold estimation. *Econometrica*, 68(3):575–603.

Javanmard, A. and Montanari, A. (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *Journal of Machine Learning Research*, 15(82):2869–2909.

Kim, J. and Pollard, D. (1990). Cube Root Asymptotics. *The Annals of Statistics*, 18(1):191 – 219.

Kock, A. B. (2016). Oracle inequalities, variable selection and uniform inference in high-dimensional correlated random effects panel data models. *Journal of Econometrics*, 195(1):71–85.

Kock, A. B. and Tang, H. (2019). Uniform inference in high-dimensional dynamic panel data models with approximately sparse fixed effects. *Econometric Theory*, 35(2):295–359.

Konishi, S. and Kitagawa, G. (1996). Generalised information criteria in model selection. *Biometrika*, 83(4):875–890.

Korovkin, V. and Makarin, A. (2023). Conflict and intergroup trade: Evidence from the 2014 russia-ukraine crisis. *American Economic Review*, 113(1):34–70.

Lee, S., Seo, M. H., and Shin, Y. (2016). The lasso for high dimensional regression with a possible change point. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(1):193–210.

Lence, S. H., Moschini, G., and Santeramo, F. G. (2018). Threshold cointegration and spatial price transmission when expectations matter. *Agricultural Economics*, 49(1):25–39.

Liefert, W. M., Mitchell, L., and Seeley, R. (2021). Economic Crises and U.S. Agricultural Exports. Economic Research Report 327195, United States Department of Agriculture, Economic Research Service.

Meinshausen, N. and Bühlmann, P. (2006). High-dimensional graphs and variable selection with the Lasso. *The Annals of Statistics*, 34(3):1436 – 1462.

Merlevède, F., Peligrad, M., and Rio, E. (2011). A Bernstein type inequality and moderate deviations for weakly dependent sequences. *Probability Theory and Related Fields volume*, 151:435–474.

Miao, K., Li, K., and Su, L. (2020a). Panel threshold models with interactive fixed effects. *Journal of Econometrics*, 219(1):137–170.

Miao, K., Su, L., and Wang, W. (2020b). Panel threshold regressions with latent group structures. *Journal of Econometrics*, 214(2):451–481.

Rebello, J. (2020). Regional integration: overview. `https://www.worldbank.org/en/topic/regional-integration/overview`.

Samuelson, P. A. (1954). The Transfer Problem and Transport Costs, II: Analysis of Effects of Trade Impediments. *The Economic Journal*, 64(254):264–289.

Semenova, V., Goldman, M., Chernozhukov, V., and Taddy, M. (2021). Estimation and inference on heterogeneous treatment effects in high-dimensional dynamic panels.

Seo, M. H. and Shin, Y. (2016). Dynamic panels with threshold effect and endogeneity. *Journal of Econometrics*, 195(2):169–186.

Tsay, R. S. (1989). Testing and modeling threshold autoregressive processes. *Journal of the American Statistical Association*, 84(405):231–240.

U.S. Department of Agriculture, Foreign Agricultural Service (2022). 2021 agricultural export yearbook. Technical report, U.S. Department of Agriculture, Foreign Agricultural Service.

van de Geer, S., Bühlmann, P., Ritov, Y., and Dezeure, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *The Annals of Statistics*, 42(3):1166 – 1202.

Varangis, P. N. and Duncan, R. C. (1993). Exchange rate pass through: An application to us and japanese steel prices. *Resources Policy*, 19(1):30–39.

Yan, H. (2023). Uniform inference in high dimensional threshold regression models. `https://hongqiangyan.github.io/files/Hongqiang_YAN_JMP.pdf`.

Yuan, M. (2010). High dimensional inverse covariance matrix estimation via linear programming. *J. Mach. Learn. Res.*, 11:2261–2286.

Zhang, C.-H. and Zhang, S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1):217–242.

APPENDICES

APPENDIX

A

CHAPTER 1

A.1 Proofs for Section 1.3

In this section of the appendix, firstly I prove the oracle inequality of prediction error. Let the event

$$\mathbb{A}_1 := \left\{ \max_{1\le j\le p} \frac{1}{n}\sum_{i=1}^{n}(X_i^{(j)})^2 \le C_2^2 + \mu\lambda \right\}$$
$$\mathbb{A}_2 := \left\{ \min_{1\le j\le p} \frac{1}{n}\sum_{i=1}^{n}(X_i^{(j)}(t_0))^2 \ge C_3^2 - \mu\lambda \right\},$$

In particular

$$\{\mathbb{A}_2\} \subseteq \left\{ \min_{1\le j\le p} \frac{1}{n}\sum_{i=1}^{n}(X_i^{(j)})^2 \ge C_3^2 - \mu\lambda \right\}$$

The following lemma provides lower bounds on the probabilities of upper bounds of second moment of regressors. B.

Lemma 3 (Probability of $\mathbb{A}_1$ and $\mathbb{A}_2$). *Let Assumption 1 be satisfied and set* λ *by* (1.3.1). *Then*

$$\mathbb{P}\{\mathbb{A}_1\} \geq 1 - \left(\frac{1}{p^{\tilde{C}_1}} + \tilde{C}_2 \frac{\frac{EM^2_{X^2}\log p}{n}}{(\log p)^2} \right)$$

$$\mathbb{P}\{\mathbb{A}_2\} \geq 1 - \left(\frac{1}{p^{\tilde{C}_3}} + \tilde{C}_4 \frac{\frac{EM^2_{X\iota_0}\log p}{n}}{(\log p)^2} \right)$$

Proof of Lemma 3. Consider the term $\|X_i^{(j)}\|_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i^{(j)})^2$. Lemma E.2(ii) of Chernozhukov et al. (2017) provides (set $\eta = 1$ and $s = 2$ in their Lemma):

$$\begin{aligned}
&\mathbb{P}\left\{ \max_{1\leq j\leq p} |\frac{1}{n}\sum_{i=1}^n (X_i^{(j)})^2 - E[(X_i^{(j)})^2]| \geq 2E\left[\max_{1\leq j\leq p}|\frac{1}{n}\sum_{i=1}^n (X_i^{(j)})^2 - E[(X_i^{(j)})^2]|\right] + \frac{t}{n}\right\} \\
\text{(A.1.1)} \qquad &\leq \exp\{-\frac{t^2}{3n\max_{1\leq j\leq p} Var[(X_i^{(j)})^2]}\} + \tilde{C}\frac{EM^2_{X^2}}{t^2}, \text{ set } t = (n\log p)^{1/2}, \text{ then I have} \\
&\leq \frac{1}{p^{\tilde{C}}} + \tilde{C}\frac{EM^2_{X^2}}{n\log p}
\end{aligned}$$

for a positive constant, $\tilde{C} > 0$. With Assumption 1, Lemma E.1 of Chernozhukov et al. (2017) provides, with $\tilde{C} > 0$ a positive constant,

$$\text{(A.1.2)} \qquad E\left[\max_{1\leq j\leq p}|\frac{1}{n}\sum_{i=1}^n (X_i^{(j)})^2 - E[(X_i^{(j)})^2]|\right] \leq \tilde{C}[\frac{\sqrt{\log p}}{\sqrt{n}} + \frac{\sqrt{EM^2_{X^2}}\log p}{n}] = O_p(\sqrt{\frac{\log p}{n}})$$

Let $c = \arg\max_{1\leq j\leq p}(\frac{1}{n}\sum_{i=1}^n (X_i^{(j)})^2 - E[(X_i^{(j)})^2])$

$$\begin{aligned}
\max_{1\leq j\leq p}(\frac{1}{n}\sum_{i=1}^n (X_i^{(j)})^2 - C_2^2) &\leq \max_{1\leq j\leq p}(\frac{1}{n}\sum_{i=1}^n (X_i^{(j)})^2 - E[(X_i^{(c)})^2]) \\
\text{(A.1.3)} \qquad &= \max_{1\leq j\leq p}(\frac{1}{n}\sum_{i=1}^n (X_i^{(j)})^2 - E[(X_i^{(j)})^2]) \\
&\leq \max_{1\leq j\leq p}|\frac{1}{n}\sum_{i=1}^n (X_i^{(j)})^2 - E[(X_i^{(j)})^2]|
\end{aligned}$$

Combine (A.1.1) with (A.1.2),

(A.1.4)

$$\mathbb{P}\left\{\max_{1\leq j\leq p}|\frac{1}{n}\sum_{i=1}^n (X_i^{(j)})^2 - E[(X_i^{(j)})^2]| \geq 2\tilde{C}[\frac{\sqrt{\log p}}{\sqrt{n}} + \frac{\sqrt{EM^2_{X^2}}\log p}{n}] + \frac{\sqrt{\log p}}{\sqrt{n}}\right\} \leq \frac{1}{p^{\tilde{C}}} + \tilde{C}\frac{EM^2_{X^2}}{n\log p}$$

To get the first part of the lemma, I combine the above display with Assumption 1 (ii), (1.3.1) and (A.1.3)

(A.1.5)
$$\begin{aligned}\mathbb{P}\{\mathbb{A}_1^c\} =&\mathbb{P}\left\{\max_{1\le j\le p}\frac{1}{n}\sum_{i=1}^{n}(X_i^{(j)})^2 - C_2^2 > C\frac{\sqrt{\log p}}{\sqrt{n}}\right\}\\ \le&\mathbb{P}\left\{\max_{1\le j\le p}|\frac{1}{n}\sum_{i=1}^{n}(X_i^{(j)})^2 - E[(X_i^{(j)})^2]| \ge C\frac{\sqrt{\log p}}{\sqrt{n}}\right\} \le \frac{1}{p^{\tilde{C}}} + \tilde{C}\frac{EM_{X^2}^2}{n\log p} = o_p(1)\end{aligned}$$

Therefore I have proved the first part of the lemma.

Next, consider $\mathbb{A}_2$. $\|(X_i^{(j)}(t_0))^2\|_n^2 = \frac{1}{n}\sum_{i=1}^{n}((X_i^{(j)}(t_0))^2)^2$. Lemma E.2(ii) of Chernozhukov et al. (2017) provides (set $\eta = 1$ and $s = 2$ in their Lemma):

(A.1.6)
$$\begin{aligned}&\mathbb{P}\left\{\max_{1\le j\le p}|\frac{1}{n}\sum_{i=1}^{n}(X_i^{(j)}(t_0))^2 - E[(X_i^{(j)}(t_0))^2]| \ge 2E\left[\max_{1\le j\le p}|\frac{1}{n}\sum_{i=1}^{n}(X_i^{(j)}(t_0))^2 - E[(X_i^{(j)}(t_0))^2]|\right] + \frac{t}{n}\right\}\\ &\le \exp\{-\frac{t^2}{3n\max_{1\le j\le p} Var[(X_i^{(j)}(t_0))^2]}\} + \tilde{C}\frac{EM_{Xt_0}^2}{t^2},\ \text{Set } t = (n\log p)^{1/2}, \text{then I have}\\ &\le \frac{1}{p^{\tilde{C}}} + \tilde{C}\frac{EM_{Xt_0}^2}{n\log p}\end{aligned}$$

for a positive constant, $\tilde{C} > 0$. With Assumption 1, Lemma E.1 of Chernozhukov et al. (2017) provides, with $\tilde{C} > 0$ a positive constant,

(A.1.7)
$$E\left[\max_{1\le j\le p}|\frac{1}{n}\sum_{i=1}^{n}(X_i^{(j)}(t_0))^2 - E[(X_i^{(j)}(t_0))^2]|\right] \le \tilde{C}[\frac{\sqrt{\log p}}{\sqrt{n}} + \frac{\sqrt{EM_{Xt_0}^2}\log p}{n}] = O_p(\sqrt{\frac{\log p}{n}})$$

Let $c = \arg\min_{1\le j\le p}(E[(X_i^{(j)}(t_0))^2] - \frac{1}{n}\sum_{i=1}^{n}(X_i^{(j)}(t_0))^2)$

(A.1.8)
$$\begin{aligned}C_3^2 - \min_{1\le j\le p}\frac{1}{n}\sum_{i=1}^{n}(X_i^{(j)}(t_0))^2 &\le \min_{1\le j\le p}(E[(X_i^{(c)}(t_0))^2] - \frac{1}{n}\sum_{i=1}^{n}(X_i^{(j)}(t_0))^2)\\ &= \min_{1\le j\le p}(E[(X_i^{(j)}(t_0))^2] - \frac{1}{n}\sum_{i=1}^{n}(X_i^{(j)}(t_0))^2)\\ &\le \max_{1\le j\le p}|\frac{1}{n}\sum_{i=1}^{n}(X_i^{(j)}(t_0))^2 - E[(X_i^{(j)}(t_0))^2]|\end{aligned}$$

Combine (A.1.6) with (A.1.7),

(A.1.9)

$$\mathbb{P}\left\{\max_{1\le j\le p}|\frac{1}{n}\sum_{i=1}^{n}(X_i^{(j)}(t_0))^2-E[(X_i^{(j)}(t_0))^2]|\ge 2\tilde{C}[\frac{\sqrt{\log p}}{\sqrt{n}}+\frac{\sqrt{EM_{Xt_0}^2}\log p}{n}]+\frac{\sqrt{\log p}}{\sqrt{n}}\right\}\le\frac{1}{p^{\tilde{C}}}+\tilde{C}\frac{EM_{Xt_0}^2}{n\log p}$$

To get the probability of event $\mathbb{A}_2$, I combine the above display with Assumption 1 (ii) , (1.3.1) and (A.1.8)

(A.1.10)

$$\begin{aligned}\mathbb{P}\{\mathbb{A}_2^c\}=&\mathbb{P}\left\{C_3^2-\min_{1\le j\le p}\frac{1}{n}\sum_{i=1}^{n}(X_i^{(j)}(t_0))^2>C\frac{\sqrt{\log p}}{\sqrt{n}}\right\}\\ \le&\mathbb{P}\left\{\max_{1\le j\le p}|\frac{1}{n}\sum_{i=1}^{n}(X_i^{(j)}(t_0))^2-E[(X_i^{(j)}(t_0))^2]|\ge C\frac{\sqrt{\log p}}{\sqrt{n}}\right\}\\ \le&\frac{1}{p^{\tilde{C}}}+\tilde{C}\frac{EM_{Xt_0}^2}{n\log p}=o_p(1)\end{aligned}$$

Therefore I have proved the lemma. □

Define the events

$$\mathbb{A}_3:=\left\{\max_{1\le j\le p}\frac{1}{\|X^{(j)}\|_n}\frac{1}{n}\sum_{i=1}^{n}U_iX_i^{(j)}\le\frac{\mu\lambda}{2}\right\},$$

$$\mathbb{A}_4:=\left\{\max_{1\le j\le p}\sup_{\tau\in\mathbb{T}}\frac{1}{\|X^{(j)}(\tau)\|_n}\frac{1}{n}\sum_{i=1}^{n}U_iX_i^{(j)}\mathbf{1}\{Q_i<\tau\}\le\frac{\mu\lambda}{2}\right\},$$

Next I provide a lower bound on the probabilities of $\mathbb{A}_1\cap\mathbb{A}_2$ with a suitable choice of λ.

Lemma 4 (Probability of $\mathbb{A}_3\cap\mathbb{A}_4$)**.** *Conditional on the events$\mathbb{A}_1\cap\mathbb{A}_2$, let Assumption 1 be satisfied and set λ by* (1.3.1). *Then*

$$\mathbb{P}\{\mathbb{A}_3\cap\mathbb{A}_4\}\ge 1-\left(\frac{1}{p^{\tilde{C}_5}}+\tilde{C}_6\frac{EM_{UX}^2}{(n\log p)}\right)-\left(\frac{1}{(pn)^{\tilde{C}_7}}+\tilde{C}_8\frac{EM_{UX}^2}{(n\log pn)}\right).$$

Proof of Lemma 4. With Assumption 1, Lemma E.1 of Chernozhukov et al. (2017) yields, with

$\tilde{C} > 0$ a positive constant,

$$\mathbb{P}\left\{\max_{1\le j\le p}|\frac{1}{n}\sum_{i=1}^{n}U_iX_i^{(j)} - E[(U_iX_i^{(j)}]| \ge 2E\left[\max_{1\le j\le p}|\frac{1}{n}\sum_{i=1}^{n}U_iX_i^{(j)} - E[(U_iX_i^{(j)}]|\right] + \frac{t}{n}\right\}$$

(A.1.11) $$\le \exp\{-\frac{t^2}{3n\max_{1\le j\le p}Var[U_iX_i^{(j)}]}\} + \tilde{C}\frac{EM_{UX}^2}{t^2},\text{ set } t=(n\log p)^{1/2}\text{,then I have}$$

$$\le \frac{1}{p^{\tilde{C}}} + \tilde{C}\frac{EM_{UX}^2}{n\log p} = o_p(1)$$

Lemma E.2(ii) of Chernozhukov et al. (2017) provides (set $\eta = 1$ and $s = 2$ in their Lemma) with $\tilde{C} > 0$ a positive constant,

(A.1.12) $$E\left[\max_{1\le j\le p}|\frac{1}{n}\sum_{i=1}^{n}U_iX_i^{(j)} - E[(U_iX_i^{(j)}]|\right] \le \tilde{C}[\frac{\sqrt{\log p}}{\sqrt{n}} + \frac{\sqrt{EM_{UX}^2}\log p}{n}] = O_p(\sqrt{\frac{\log p}{n}})$$

Using that I are on the set on $\mathbb{A}_1$,

(A.1.13)
$$\max_{1\le j\le p}\frac{1}{\|X^{(j)}\|_n}\frac{1}{n}\sum_{i=1}^{n}U_iX_i^{(j)} \le \frac{1}{\min_{1\le j\le p}\|X^{(j)}\|_n}\max_{1\le j\le p}\frac{1}{n}\sum_{i=1}^{n}U_iX_i^{(j)} \le \frac{1}{\sqrt{C_3^2-\mu\lambda}}\max_{1\le j\le p}\frac{1}{n}\sum_{i=1}^{n}U_iX_i^{(j)}$$

Combine (A.1.11) with (A.1.12),

$$\mathbb{P}\left\{\max_{1\le j\le p}|\frac{1}{n}\sum_{i=1}^{n}U_iX_i^{(j)}| \ge 2\tilde{C}[\frac{\sqrt{\log p}}{\sqrt{n}} + \frac{(EM_{UX}^2)^{1/2}\log p}{n}] + \frac{\sqrt{\log p}}{\sqrt{n}}\right\} \le \frac{1}{p^{\tilde{C}}} + \tilde{C}\frac{EM_{UX}^2}{n\log p} = o_p(1),$$

Then,

$$\begin{aligned}\mathbb{P}\{\mathbb{A}_3^c\} =& \mathbb{P}\left\{\max_{1\le j\le p}\frac{1}{\|X^{(j)}\|_n}\frac{1}{n}\sum_{i=1}^{n}U_iX_i^{(j)} > \frac{\mu C}{2}\frac{\sqrt{\log p}}{\sqrt{n}}\right\}\\ \le& \mathbb{P}\left\{\max_{1\le j\le p}\frac{1}{n}\sum_{i=1}^{n}U_iX_i^{(j)} \ge \frac{\mu C}{2}\sqrt{C_3^2-\mu\lambda}\frac{\sqrt{\log p}}{\sqrt{n}}\right\}\\ \le& \mathbb{P}\left\{\max_{1\le j\le p}\frac{1}{n}\sum_{i=1}^{n}U_iX_i^{(j)} \ge \tilde{C}\frac{\sqrt{\log p}}{\sqrt{n}}\right\} + o_p(1)\\ \le& \frac{1}{p^{\tilde{C}}} + \tilde{C}\frac{EM_{UX}^2}{n\log p} + o_p(1) = o_p(1)\end{aligned}$$

This shows also that

$$\max_{1\le j\le p}\left|\frac{1}{\|X^{(j)}\|_n}\frac{1}{n}\sum_{i=1}^{n}U_iX_i^{(j)}\right| = O_p\left(\frac{\sqrt{\log p}}{\sqrt{n}}\right) \tag{A.1.14}$$

Next, consider the event $\mathbb{A}_4$. To show the sup norm over τ, I adapt the proof of equitation (A.1) and (A.2) in Lemma A.1 of Callot et al. (2017) to my purpose. Conditional on $\mathbb{A}_4$, then sort $\{X_i, U_i, Q_i\}_{i=1}^n$ by $(Q_1, \cdots Q_n)$ in ascending order, for $j = 1, \cdots p$,

$$\begin{aligned}
&\mathbb{P}\left\{\max_{1\le j\le p}\sup_{\tau\in\mathbb{T}}\frac{1}{\|X^{(j)}(\tau)\|_n}\left|\frac{1}{n}\sum_{i=1}^{n}U_iX_i^{(j)}(\tau)\right|\ge t\right\}\\
\le&\mathbb{P}\left\{\frac{1}{\min_{1\le j\le p}\|X^{(j)}(t_0)\|_n}\max_{1\le j\le p}\sup_{\tau\in\mathbb{T}}\left|\frac{1}{n}\sum_{i=1}^{n}U_iX_i^{(j)}(\tau)\right|\ge t\right\}\\
=&\mathbb{P}\left\{\max_{1\le j\le p}\max_{1\le k\le n}\left|\frac{1}{n}\sum_{i=1}^{k}U_iX_i^{(j)}\right|\ge t\min_{1\le j\le p}\left\|X^{(j)}(t_0)\right\|_n\right\}\\
\le&\mathbb{P}\left\{\max_{1\le j\le p}\max_{1\le k\le n}\left|\frac{1}{n}\sum_{i=1}^{k}U_iX_i^{(j)}\right|\ge t\sqrt{C_3^2-\mu\lambda}\right\}
\end{aligned}$$

Denote a deterministic upper triangular matrix with all ones

$$\Xi_{n,n} = \begin{pmatrix} 1 & 1 & \cdots & 1\\ 0 & 1 & \cdots & 1\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & 1\end{pmatrix}, \tag{A.1.15}$$

and let $\xi_i^{(k)}$ is the $i-th$ row, $k-th$ column element of $\Xi_{n,n}$, then

$$\max_{1\le j\le p}\max_{1\le k\le n}\left|\sum_{i=1}^{k}U_iX_i^{(j)}\right| = \max_{1\le j\le p}\max_{1\le k\le n}\left|\sum_{i=1}^{n}U_iX_i^{(j)}\xi_i^{(k)}\right|$$

$U_iX_i^{(j)}\xi_i^{(k)}$ is independent centered random variable(not identical),

$$\max_{1\le j\le p}\max_{1\le k\le n}\left|\sum_{i=1}^{k}U_iX_i^{(j)}\right| = \max_{1\le j\le p}\max_{1\le k\le n}\left|\sum_{i=1}^{n}U_iX_i^{(j)}\xi_i^{(k)}\right|$$

$U_iX_i^{(j)}\xi_i^{(k)}$ is independent centered random variable(not identical),

$$\frac{\max_{1\le j\le p}\max_{1\le k\le n}\sum_{i=1}^{n} Var[U_iX_i^{(j)}\xi_i^{(k)}]}{n} = \frac{\max_{1\le j\le p}\max_{1\le k\le n}\sum_{i=1}^{k} Var[U_iX_i^{(j)}]}{n} \le \frac{\max_{1\le j\le p}\sum_{i=1}^{n} Var[U_iX_i^{(j)}]}{n} = \max_{1\le j\le p} Var[U_iX_i^{(j)}] < \infty$$

and $\max_{1\le j\le p}\max_{1\le i\le k\le n}|U_iX_i^{(j)}\xi_i^{(k)}| < M_{UX}$. So under Assumption 1, conditions for maximal inequalities are stratified automatically. Lemma E.2(ii) of Chernozhukov et al. (2017) provides (set $\eta = 1$ and $s = 2$ in their Lemma) with $\tilde{C} > 0$ a positive constant,

(A.1.16)

$$\mathbb{P}\left\{\max_{1\le j\le p}\max_{1\le k\le n}|\sum_{i=1}^{n} U_iX_i^{(j)}\xi_i^{(k)}| \ge 2E\left[\max_{1\le j\le p}\max_{1\le k\le n}|\sum_{i=1}^{n} U_iX_i^{(j)}\xi_i^{(k)}|\right] + t\right\}$$
$$\le \exp\{-\frac{t^2}{3\max_{1\le j\le p}\max_{1\le k\le n}\sum_{i=1}^{n} Var[U_iX_i^{(j)}\xi_i^{(k)}]}\} + \tilde{C}\frac{EM_{UX}^2}{t^2},\ \text{set } t = (n\log pn)^{1/2}, \text{then I have}$$
$$\le \frac{1}{(pn)^{\tilde{C}}} + \tilde{C}\frac{EM_{UX}^2}{(n\log pn)}$$

With Assumption 1, Lemma E.1 of Chernozhukov et al. (2017) yields, with $\tilde{C} > 0$ a positive constant,

$$(\text{A.1.17})\quad \begin{aligned} &E\left[\max_{1\le j\le p}\max_{1\le k\le n}|\sum_{i=1}^{n} U_iX_i^{(j)}\xi_i^{(k)}|\right] \\ \le&\sqrt{\max_{1\le j\le p}\max_{1\le k\le n}\sum_{i=1}^{n} Var[U_iX_i^{(j)}\xi_i^{(k)}]}\sqrt{\log pn} + \sqrt{EM_{UX}^2}\log pn \\ \le&\tilde{C}\sqrt{n\log(pn)} + \sqrt{EM_{UX}^2}\log pn = O_p(\sqrt{n\log(pn)}) \end{aligned}$$

Combine (A.1.16) with (A.1.17) and I consider $p > n$,

$$\begin{aligned} &\mathbb{P}\left\{\max_{1\le j\le p}\max_{1\le k\le n}|\frac{1}{n}\sum_{i=1}^{n} U_iX_i^{(j)}\xi_i^{(k)}| \ge 2\tilde{C}[\frac{\sqrt{\log(pn)}}{\sqrt{n}} + \frac{\sqrt{EM_{UX}^2}\log(pn)}{n}] + \frac{\sqrt{\log(pn)}}{\sqrt{n}}\right\} \\ \le&\mathbb{P}\left\{\max_{1\le j\le p}\max_{1\le k\le n}|\frac{1}{n}\sum_{i=1}^{n} U_iX_i^{(j)}\xi_i^{(k)}| \ge 2\tilde{C}[\frac{\sqrt{\log p}}{\sqrt{n}} + \frac{\sqrt{EM_{UX}^2}\log p}{n}] + \frac{\sqrt{\log p}}{\sqrt{n}}\right\} \\ \le&\frac{1}{(pn)^{\tilde{C}}} + \tilde{C}\frac{EM_{UX}^2}{(n\log(pn))}, \end{aligned}$$

Taking expectations over $(Q_1, \cdots Q_n)$ and set λ by (1.3.1) yields,

$$\begin{aligned}
\mathbb{P}\{\mathbb{A}_4^c\} &= \mathbb{P}\left\{ \max_{1\le j\le p} \sup_{\tau\in\mathbb{T}} \frac{1}{\|X^{(j)}(\tau)\|_n} \frac{1}{n}\sum_{i=1}^n U_i X_i^{(j)} \mathbf{1}\{Q_i<\tau\} \ge \frac{\mu C}{2}\frac{\sqrt{\log p}}{\sqrt{n}} \right\} \\
&\le \mathbb{P}\left\{ \max_{1\le j\le p} \sup_{\tau\in\mathbb{T}} \frac{1}{n}\sum_{i=1}^n U_i X_i^{(j)} \mathbf{1}\{Q_i<\tau\} \ge \frac{\mu C}{2}\sqrt{C_3^2-\mu\lambda}\frac{\sqrt{\log p}}{\sqrt{n}} \right\} \qquad \text{(A.1.18)} \\
&\le \frac{1}{(pn)^{\tilde{C}}} + \tilde{C}\frac{EM_{UX}^2}{(n\log pn)} = o_p(1)
\end{aligned}$$

This shows that

$$\left| \max_{1\le j\le p} \sup_{\tau\in\mathbb{T}} \frac{1}{\|X^{(j)}(\tau)\|_n} \frac{1}{n}\sum_{i=1}^n U_i X_i^{(j)} \mathbf{1}\{Q_i<\tau\} \right| = O_p\left(\frac{\sqrt{\log p}}{\sqrt{n}}\right) \qquad \text{(A.1.19)}$$

Since $\mathbb{P}\{\mathbb{A}_3 \cap \mathbb{A}_4\} \ge 1 - \mathbb{P}\{\mathbb{A}_3^c\} - \mathbb{P}\{\mathbb{A}_4^c\}$, I have proved the lemma. □

Define $J_0 := J(\alpha_0)$, $\widehat{\boldsymbol{D}} = \widehat{\boldsymbol{D}}(\hat{\tau})$, $\boldsymbol{D} = \boldsymbol{D}(\tau_0)$ and $R_n := R_n(\alpha_0, \tau_0)$, where

$$R_n(\alpha,\tau) := 2n^{-1}\sum_{i=1}^n U_i X_i'\delta\left\{1(Q_i<\widehat{\tau}) - 1(Q_i<\tau)\right\}.$$

Lemma 5. *Conditional on the events $\mathbb{A}_1$, $\mathbb{A}_2$, $\mathbb{A}_3$ and $\mathbb{A}_4$, for $0<\mu<1$ I have*

$$\left\|\widehat{f} - f_0\right\|_n^2 + (1-\mu)\lambda\left\|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)\right\|_1 \le 2\lambda\left\|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)_{J_0}\right\|_1 + \left\|f_{(\alpha_0,\widehat{\tau})} - f_0\right\|_n^2. \qquad \text{(A.1.20)}$$

Proof of Lemma 5. We begin by noting that (1.2.4)

$$\widehat{S}_n + \lambda\left\|\widehat{\boldsymbol{D}}\widehat{\alpha}\right\|_1 \le S_n(\alpha,\tau) + \lambda\|\boldsymbol{D}(\tau)\alpha\|_1 \qquad \text{(A.1.21)}$$

$$\widehat{S}_n - S_n(\alpha,\tau) \le \lambda\|\boldsymbol{D}(\tau)\alpha\|_1 - \lambda\left\|\widehat{\boldsymbol{D}}\widehat{\alpha}\right\|_1 \qquad \text{(A.1.22)}$$

for all $(\alpha,\tau)\in\mathbb{R}^{2p}\times\mathbb{T}$. Inserting (1.2.3) to left side of (A.1.22)

$$
\begin{aligned}
&\widehat{S}_n - S_n(\alpha,\tau)\\
&= n^{-1}\|\boldsymbol{y}-\boldsymbol{X}(\widehat{\tau})\widehat{\alpha}\|_2^2 - n^{-1}\|\boldsymbol{y}-\boldsymbol{X}(\tau)\alpha\|_2^2\\
&= n^{-1}\sum_{i=1}^n\left[U_i-\left(\boldsymbol{X}_i(\widehat{\tau})'\widehat{\alpha}-\boldsymbol{X}_i(\tau_0)'\alpha_0\right)\right]^2 - n^{-1}\sum_{i=1}^n\left[U_i-\left(\boldsymbol{X}_i(\tau)'\alpha-\boldsymbol{X}_i(\tau_0)'\alpha_0\right)\right]^2\\
&= n^{-1}\sum_{i=1}^n\left\{\boldsymbol{X}_i(\widehat{\tau})'\widehat{\alpha}-\boldsymbol{X}_i(\tau_0)'\alpha_0\right\}^2 - n^{-1}\sum_{i=1}^n\left\{\boldsymbol{X}_i(\tau)'\alpha-\boldsymbol{X}_i(\tau_0)'\alpha_0\right\}^2\\
&\quad -2n^{-1}\sum_{i=1}^n U_i\left\{\boldsymbol{X}_i(\widehat{\tau})'\widehat{\alpha}-\boldsymbol{X}_i(\tau)'\alpha\right\}\\
&= \left\|\widehat{f}-f_0\right\|_n^2 - \left\|f_{(\alpha,\tau)}-f_0\right\|_n^2\\
&\quad -2n^{-1}\sum_{i=1}^n U_iX_i'(\widehat{\beta}-\beta) - 2n^{-1}\sum_{i=1}^n U_i\left\{X_i'\widehat{\delta}1(Q_i<\widehat{\tau})-X_i'\delta 1(Q_i<\tau)\right\}.
\end{aligned}
$$

Further, the last term on the right side of above can be written as

$$
\begin{aligned}
&n^{-1}\sum_{i=1}^n U_i\left\{X_i'\widehat{\delta}1(Q_i<\widehat{\tau})-X_i'\delta 1(Q_i<\tau)\right\}\\
&= n^{-1}\sum_{i=1}^n U_iX_i'(\widehat{\delta}-\delta)1(Q_i<\widehat{\tau}) + n^{-1}\sum_{i=1}^n U_iX_i'\delta\left\{1(Q_i<\widehat{\tau})-1(Q_i<\tau)\right\}.
\end{aligned}
$$

Then, (A.1.22) can be bounded as follows:

$$
\begin{aligned}
\left\|\widehat{f}-f_0\right\|_n^2 \le{}& \left\|f_{(\alpha,\tau)}-f_0\right\|_n^2 + \lambda\|\boldsymbol{D}(\tau)\alpha\|_1 - \lambda\left\|\widehat{\boldsymbol{D}}\widehat{\alpha}\right\|_1\\
&+2n^{-1}\sum_{i=1}^n U_iX_i'(\widehat{\beta}-\beta) + 2n^{-1}\sum_{i=1}^n U_iX_i'(\widehat{\delta}-\delta)1(Q_i<\widehat{\tau})\\
&+2n^{-1}\sum_{i=1}^n U_iX_i'\delta\left\{1(Q_i<\widehat{\tau})-1(Q_i<\tau)\right\}.
\end{aligned}
$$

Conditional on the events $\mathbb{A}_1$, $\mathbb{A}_2$, $\mathbb{A}_3$ and $\mathbb{A}_4$, I have

$$
\begin{aligned}
\left\|\widehat{f}-f_0\right\|_n^2 \le{}& \left\|f_{(\alpha,\tau)}-f_0\right\|_n^2 + \mu\lambda\sum_{j=1}^p\left\|X^{(j)}\right\|_n(\widehat{\beta_j}-\beta_j) + \mu\lambda\sum_{j=1}^p\left\|X^{(j)}(\widehat{\tau})\right\|_n(\widehat{\delta_j}-\delta_j)\\
&+\lambda\|\boldsymbol{D}(\tau)\alpha\|_1 - \lambda\left\|\widehat{\boldsymbol{D}}\widehat{\alpha}\right\|_1 + R_n(\alpha,\tau)\\
\le{}& \left\|f_{(\alpha,\tau)}-f_0\right\|_n^2 + \mu\lambda\left\|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha)\right\|_1 + \lambda\|\boldsymbol{D}(\tau)\alpha\|_1 - \lambda\left\|\widehat{\boldsymbol{D}}\widehat{\alpha}\right\|_1 + R_n(\alpha,\tau)
\end{aligned}
\tag{A.1.23}
$$

for all $(\alpha,\tau)\in\mathbb{R}^{2p}\times\mathbb{T}$. Note the fact that

$$\left|\widehat{\alpha}^{(j)}-\alpha_0^{(j)}\right|+\left|\alpha_0^{(j)}\right|-\left|\widehat{\alpha}^{(j)}\right|=0 \text{ for } j\notin J_0 \tag{A.1.24}$$

$$\left\|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)\right\|_1=\left\|\left[\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)\right]_{J_0}\right\|_1+\left\|\left[\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)\right]_{J_0^C}\right\|_1=\left\|\left[\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)\right]_{J_0}\right\|_1+\left\|\left[\widehat{\boldsymbol{D}}\widehat{\alpha}\right]_{J_0^C}\right\|_1, \tag{A.1.25}$$

(A.1.26)

$$\begin{aligned}
\|\boldsymbol{D}\alpha_0\|_1-\left\|\widehat{\boldsymbol{D}}\widehat{\alpha}\right\|_1&=\left\|[\boldsymbol{D}\alpha_0]_{J_0}\right\|_1-\left\|\left[\widehat{\boldsymbol{D}}\widehat{\alpha}\right]_{J_0}\right\|_1-\left\|\left[\widehat{\boldsymbol{D}}\widehat{\alpha}\right]_{J_0^C}\right\|_1\\
&=\left\|[\boldsymbol{D}\alpha_0]_{J_0}\right\|_1-\left\|\left[\widehat{\boldsymbol{D}}\alpha_0\right]_{J_0}\right\|_1+\left\|\left[\widehat{\boldsymbol{D}}\alpha_0\right]_{J_0}\right\|_1-\left\|\left[\widehat{\boldsymbol{D}}\widehat{\alpha}\right]_{J_0}\right\|_1-\left\|\left[\widehat{\boldsymbol{D}}\widehat{\alpha}\right]_{J_0^C}\right\|_1\\
\text{using triangle inequality}&\le\left|\left\|[\boldsymbol{D}\alpha_0]_{J_0}\right\|_1-\left\|\left[\widehat{\boldsymbol{D}}\alpha_0\right]_{J_0}\right\|_1\right|+\left\|\left[\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)\right]_{J_0}\right\|_1-\left\|\left[\widehat{\boldsymbol{D}}\widehat{\alpha}\right]_{J_0^C}\right\|_1\\
&=\left|\|\boldsymbol{D}\alpha_0\|_1-\left\|\widehat{\boldsymbol{D}}\alpha_0\right\|_1\right|+\left\|\left[\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)\right]_{J_0}\right\|_1-\left\|\left[\widehat{\boldsymbol{D}}\widehat{\alpha}\right]_{J_0^C}\right\|_1.
\end{aligned}$$

Consider (A.1.20), conditional on the events $\mathbb{A}_1$, $\mathbb{A}_2$, $\mathbb{A}_3$ and $\mathbb{A}_4$, add $(1-\mu)\lambda\left\|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha)\right\|_1$ on both sides of (A.1.23)(evaluating at $(\alpha,\tau)=(\alpha_0,\widehat{\tau})$, $R_n(\alpha_0,\widehat{\tau})=0$),to get

$$\begin{aligned}
&\|\widehat{f}-f_0\|_n^2+(1-\mu)\lambda\left\|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha)\right\|_1\\
&\le\lambda\left(\left\|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)\right\|_1+\left\|\widehat{\boldsymbol{D}}\alpha_0\right\|_1-\left\|\widehat{\boldsymbol{D}}\widehat{\alpha}\right\|_1\right)+\left\|f_{(\alpha_0,\widehat{\tau})}-f_0\right\|_n^2\\
&\quad\text{using (A.1.25) and (A.1.26)}\\
&\le\lambda\left(\left\|\left[\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)\right]_{J_0}\right\|_1+\left\|\left[\widehat{\boldsymbol{D}}\widehat{\alpha}\right]_{J_0^C}\right\|_1+\left|\|\boldsymbol{D}\alpha_0\|_1-\left\|\widehat{\boldsymbol{D}}\alpha_0\right\|_1\right|+\left\|\left[\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)\right]_{J_0}\right\|_1-\left\|\left[\widehat{\boldsymbol{D}}\widehat{\alpha}\right]_{J_0^C}\right\|_1\right)+\left\|f_{(\alpha_0,\widehat{\tau})}-f_0\right\|_n^2\\
&\le2\lambda\left\|\left[\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)\right]_{J_0}\right\|_1+\left\|f_{(\alpha_0,\widehat{\tau})}-f_0\right\|_n^2,
\end{aligned}$$

which proves $(A.1.20)$. □

We are ready to establish the prediction consistency of the LASSO estimator.

Proof of Lemma 1. Conditional on the events $\mathbb{A}_1$, $\mathbb{A}_2$, $\mathbb{A}_3$ and $\mathbb{A}_4$, add $(1-\mu)\lambda\left\|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha)\right\|_1$ on both sides of (A.1.23) (evaluating at $(\alpha,\tau)=(\alpha_0,\tau_0)$, $\widehat{f}(\alpha_0,\widehat{\tau})-f_0=0$),to get

(A.1.27)
$$
\begin{aligned}
&\left\|\widehat{f}-f_0\right\|_n^2+(1-\mu)\lambda\left\|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)\right\|_1\\
\leq&\lambda\left(\left\|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)\right\|_1+\left\|\widehat{\boldsymbol{D}}\alpha_0\right\|_1-\left\|\widehat{\boldsymbol{D}}\widehat{\alpha}\right\|_1\right)+R_n\\
&\text{using (A.1.25) and (A.1.26)}\\
\leq&\lambda\left(\left\|\left[\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)\right]_{J_0}\right\|_1+\left\|\left[\widehat{\boldsymbol{D}}\widehat{\alpha}\right]_{J_0^C}\right\|_1+\left|\|\boldsymbol{D}\alpha_0\|_1-\left\|\widehat{\boldsymbol{D}}\alpha_0\right\|_1\right|+\left\|\left[\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)\right]_{J_0}\right\|_1-\left\|\left[\widehat{\boldsymbol{D}}\widehat{\alpha}\right]_{J_0^C}\right\|_1\right)+R_n\\
\leq&2\lambda\left\|\left[\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)\right]_{J_0}\right\|_1+\lambda\left|\|\widehat{\boldsymbol{D}}\alpha_0\|_1-\|\boldsymbol{D}\alpha_0\|_1\right|+R_n,
\end{aligned}
$$

The 3 terms on the right side can be bounded as follows using Hölder's inequality :

$$
|R_n|\leq 2\mu\lambda\sum_{j=1}^{p}\left\|X^{(j)}\right\|_n|\delta_0^{(j)}|\leq 2\mu\|\delta_0\|_1\lambda\sqrt{C_2^2+\mu\lambda}., \tag{A.1.28}
$$

$$
\left\|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)_{J_0}\right\|_1\leq\left\|\widehat{\boldsymbol{D}}\right\|_\infty\left\|(\widehat{\alpha}-\alpha_0)_{J_0}\right\|_1\leq\left\|(\widehat{\alpha}-\alpha_0)_{J_0}\right\|_1\sqrt{C_2^2+\mu\lambda}, \tag{A.1.29}
$$

$$
\left|\|\widehat{\boldsymbol{D}}\alpha_0\|_1-\|\boldsymbol{D}\alpha_0\|_1\right|\leq\left\|(\widehat{\boldsymbol{D}}-\boldsymbol{D})\alpha_0\right\|_1\leq\left\|\widehat{\boldsymbol{D}}-\boldsymbol{D}\right\|_\infty\|\alpha_0\|_1\leq 2\|\alpha_0\|_1\sqrt{C_2^2+\mu\lambda} \tag{A.1.30}
$$

Combine (A.1.28), (A.1.29) and (A.1.30) with (A.1.27) yields

$$
\begin{aligned}
\left\|\widehat{f}-f_0\right\|_n^2\leq&\left(2\left\|(\widehat{\alpha}-\alpha_0)_{J_0}\right\|_1+2\|\alpha_0\|_1+2\mu\|\delta_0\|_1\right)\lambda\left(C_2^2+\mu\lambda\right)^{\frac{1}{2}}\\
\leq&(6+2\mu)C_1\left(C_2^2+\mu\lambda\right)^{\frac{1}{2}}s_0\lambda
\end{aligned}
$$

which is (1.3.2). □

A.2 Proofs for Section 1.3.1

My first result is a preliminary lemma that can be used to prove adaptive restricted eigenvalue condition.

Lemma 6. *Let Assumptions 1 hold, for a universal constant $C>0$,*

$$
\left\|\frac{1}{n}X_i'X_i-E\left[X_i'X_i\right]\right\|_\infty\leq C\frac{\sqrt{\log p}}{\sqrt{n}}
$$

with probability at least $1-\left(\frac{1}{p^{2\tilde{C}}}+\tilde{C}\frac{EM_{XX}^2}{n\log p^2}\right)$,

$$\sup_{\tau\in\mathbb{T}}\left\|\frac{1}{n}X_i(\tau)'X_i(\tau)-E\left[X_i(\tau)'X_i(\tau)\right]\right\|_\infty\leq C\frac{\sqrt{\log p}}{\sqrt{n}}$$

with probability at least $1-\left(\frac{1}{(p^2n)^{\tilde{C}}}+\tilde{C}\frac{EM_{XX}^2}{(n\log p^2 n)}\right)$.

Proof. With Assumption 1, Lemma E.2(ii) of Chernozhukov et al. (2017) with $\tilde{C}>0$ a positive constant provides (set $\eta=1$ and $s=2$ in their Lemma) :

(A.2.1)

$$\mathbb{P}\left\{\max_{1\leq j,l\leq p}|\frac{1}{n}\sum_{i=1}^n X_i^{(j)}X_i^{(l)}-E[(X_i^{(j)}X_i^{(l)}]|\geq 2E\left[\max_{1\leq j,l\leq p}|\frac{1}{n}\sum_{i=1}^n X_i^{(j)}X_i^{(l)}-E[(X_i^{(j)}X_i^{(l)}]|\right]+\frac{t}{n}\right\}$$
$$\leq\exp\{-\frac{t^2}{3n\max_{1\leq j\leq p}Var[X_i^{(l)}X_i^{(j)}]}\}+\tilde{C}\frac{EM_{XX}^2}{t^2},\text{ set } t=(n\log p^2)^{1/2}\text{,then I have}$$
$$\leq\frac{1}{p^{2\tilde{C}}}+\tilde{C}\frac{EM_{XX}^2}{n\log p^2}=o_p(1)$$

Lemma E.1 of Chernozhukov et al. (2017) with $\tilde{C}>0$ a positive constant yields:

$$\text{(A.2.2)}\quad E\left[\max_{1\leq j,l\leq p}|\frac{1}{n}\sum_{i=1}^n X_i^{(j)}X_i^{(j)}-E[(X_i^{(j)}X_i^{(j)}]\right]\leq\tilde{C}[\frac{\sqrt{\log p^2}}{\sqrt{n}}+\frac{\sqrt{EM_{XX}^2}\log p^2}{n}]=O_p(\sqrt{\frac{\log p}{n}}),$$

with $\tilde{C}>0$ a positive constant. Combine (A.2.1) with (A.2.2),

$$\mathbb{P}\left\{\max_{1\leq j\leq p}\max_{1\leq l\leq p}|\frac{1}{n}\sum_{i=1}^n X_i^{(j)}X_i^{(l)}-E[X_i^{(j)}X_i^{(l)}]|\geq 2\tilde{C}[\frac{\sqrt{\log p^2}}{\sqrt{n}}+\frac{\sqrt{EM_{XX}^2}\log p^2}{n}]+\frac{\sqrt{\log p^2}}{\sqrt{n}}\right\}$$
$$=\mathbb{P}\left\{\max_{1\leq j\leq p}\max_{1\leq l\leq p}|\frac{1}{n}\sum_{i=1}^n X_i^{(j)}X_i^{(l)}-E[X_i^{(j)}X_i^{(l)}]|\geq\left[2\tilde{C}\left(1+\frac{(EM_{XX}^2)^{1/2}\sqrt{\log p^2}}{\sqrt{n}}\right)+1\right]\frac{\sqrt{\log p^2}}{\sqrt{n}}\right\}$$
$$\leq\frac{1}{p^{2\tilde{C}}}+\tilde{C}\frac{EM_{XX}^2}{n\log p^2}=o_p(1),$$

This shows that

$$\max_{1\leq j\leq p}\max_{1\leq l\leq p}|\frac{1}{n}\sum_{i=1}^n X_i^{(j)}X_i^{(l)}-E[X_i^{(j)}X_i^{(l)}]|=O_p(\frac{\sqrt{\log p}}{\sqrt{n}})$$

Next, to show the sup norm over τ, I adapt the proof of equitation (A.1) in Lemma A.1 of

Callot et al. (2017) to my purpose.

Sort (X_i, U_i, Q_i) $i = \{1 \cdots n\}$ by $(Q_1, \cdots Q_n)$ in ascending order, then

$$\begin{aligned} &\mathbb{P}\left\{\max_{1\le j,l\le p}\sup_{\tau\in\mathbb{T}}|\frac{1}{n}\sum_{i=1}^{n}\left(X_i^{(j)}X_i^{(l)}1(Q_i<\tau)-1(Q_i<\tau)E[X_i^{(j)}X_i^{(l)}]\right)|\ge t\right\} \\ &\le\mathbb{P}\left\{\max_{1\le j,l\le p}\max_{1\le k\le n}|\frac{1}{n}\sum_{i=1}^{k}\left(X_i^{(j)}X_i^{(l)}-E[X_i^{(j)}X_i^{(l)}]\right)|\ge t\right\} \end{aligned} \tag{A.2.3}$$

Recall matrix $\Xi_{n,n}$ in (A.1.15), and $\xi_i^{(k)}$ which is the $i-th$ row, $k-th$ column element of it, then

$$\max_{1\le j,l\le p}\max_{1\le k\le n}|\frac{1}{n}\sum_{i=1}^{k}\left(X_i^{(j)}X_i^{(l)}-E[X_i^{(j)}X_i^{(l)}]\right)|=\max_{1\le j,l\le p}\max_{1\le k\le n}|\frac{1}{n}\sum_{i=1}^{n}\left(X_i^{(j)}X_i^{(l)}-E[X_i^{(j)}X_i^{(l)}]\right)\xi_i^{(k)}|$$

$\left(X_i^{(j)}X_i^{(l)}-E[X_i^{(j)}X_i^{(l)}]\right)\xi_i^{(k)}$ is independent centered random variable (not identical) across i,

$$\begin{aligned} &\frac{\max_{1\le j,l\le p}\max_{1\le k\le n}\sum_{i=1}^{n} var[\left(X_i^{(j)}X_i^{(l)}-E[X_i^{(j)}X_i^{(l)}]\right)\xi_i^{(k)}]}{n} \\ &\le\frac{\max_{1\le j,l\le p}\sum_{i=1}^{n} var[\left(X_i^{(j)}X_i^{(l)}-E[X_i^{(j)}X_i^{(l)}]\right)]}{n}<\infty, \end{aligned}$$

$$\begin{aligned} &\max_{1\le j,l\le p}\max_{1\le i,k\le n}|\left(X_i^{(j)}X_i^{(l)}-E[X_i^{(j)}X_i^{(l)}]\right)\xi_i^{(k)}| \\ &\le\max_{1\le j,l\le p}\max_{1\le i\le n}|X_i^{(j)}X_i^{(l)}-E[X_i^{(j)}X_i^{(l)}]|=M_{XX}. \end{aligned}$$

So under assumption 1, conditions for maximal inequalities are stratified automatically. We can apply a 3-layer Maximal Inequalities over j,l,k. Lemma E.2(ii) of Chernozhukov et al. (2017) (set $\eta=1$ and $s=2$ in their Lemma) with $\tilde{C}>0$ a positive constant yields:

(A.2.4)

$$\begin{aligned} &\mathbb{P}\left\{\max_{1\le j,l\le p}\max_{1\le k\le n}|\sum_{i=1}^{n}\left(X_i^{(j)}X_i^{(l)}-E[X_i^{(j)}X_i^{(l)}]\right)\xi_i^{(k)}|\ge 2E\left[\max_{1\le j,l\le p}\max_{1\le k\le n}|\sum_{i=1}^{n}\left(X_i^{(j)}X_i^{(l)}-E[X_i^{(j)}X_i^{(l)}]\right)\xi_i^{(k)}|\right]+t\right\} \\ &\le\exp\{-\frac{t^2}{3\max_{1\le j,l\le p}\sum_{i=1}^{n} var[\left(X_i^{(j)}X_i^{(l)}-E[X_i^{(j)}X_i^{(l)}]\right)\xi_i^{(k)}]}\}+\tilde{C}\frac{EM_{XX}^2}{t^2},\ \text{set } t=[n\log(p^2n)]^{1/2}\text{,then I have} \\ &\le\frac{1}{(p^2n)^{\tilde{C}}}+\tilde{C}\frac{EM_{XX}^2}{(n\log(p^2n))} \end{aligned}$$

Lemma E.1 of Chernozhukov et al. (2017) provides:

(A.2.5)

$$E\left[\max_{1\le j,l\le p}\max_{1\le k\le n}|\sum_{i=1}^{n}\left(X_i^{(j)}X_i^{(l)}-E[X_i^{(j)}X_i^{(l)}]\right)\xi_i^{(k)}|\right]\le\tilde{C}(\sqrt{n}\sqrt{\log(p^2n)}+\sqrt{EM_{XX}^2}\log(p^2n))=O_p(\sqrt{n\log(p^2n)})$$

Combine (A.2.4) with (A.2.5),

$$\mathbb{P}\left\{\max_{1\le j,l\le p}\max_{1\le k\le n}|\frac{1}{n}\sum_{i=1}^{n}\left(X_i^{(j)}X_i^{(l)}-E[X_i^{(j)}X_i^{(l)}]\right)\xi_i^{(k)}|\ge 2\tilde{C}[\frac{\sqrt{\log(p^2n)}}{\sqrt{n}}+\frac{\sqrt{EM_{XX}^2}\log(p^2n)}{n}]+\frac{\sqrt{\log(p^2n)}}{\sqrt{n}}\right\}$$
$$\le\mathbb{P}\left\{\max_{1\le j,l\le p}\max_{1\le k\le n}|\frac{1}{n}\sum_{i=1}^{n}\left(X_i^{(j)}X_i^{(l)}-E[X_i^{(j)}X_i^{(l)}]\right)\xi_i^{(k)}|\ge\left[2\tilde{C}\left(1+\frac{(EM_{XX}^2)^{1/2}\sqrt{\log(p^2n)}}{\sqrt{n}}\right)+1\right]\frac{\sqrt{\log(p^2n)}}{\sqrt{n}}\right\}$$
$$\le\frac{1}{(p^2n)^{\tilde{C}}}+\tilde{C}\frac{EM_{XX}^2}{(n\log(p^2n))},$$

Then plug-in $t=\left[2\tilde{C}\left(1+\frac{(EM_{XX}^2)^{1/2}\sqrt{\log(p^2n)}}{\sqrt{n}}\right)+1\right]\frac{\sqrt{\log(p^2n)}}{\sqrt{n}}$ in(A.2.3) and take expectations over $(Q_1,\cdots,Q_n)\in(0,1)$ yields,

(A.2.6)

$$\mathbb{P}\{\max_{1\le j,l\le p}\sup_{\tau\in\mathbb{T}}|\frac{1}{n}\sum_{i=1}^{n}X_i^{(j)}X_i^{(l)}1(Q_i<\tau)-E[1(Q_i<\tau)]E[X_i^{(j)}X_i^{(l)}]|$$
$$\ge\left[2\tilde{C}\left(1+\frac{(EM_{XX}^2)^{1/2}\sqrt{\log(p^2n)}}{\sqrt{n}}\right)+1\right]\frac{\sqrt{\log(p^2n)}}{\sqrt{n}}\}$$
$$=\mathbb{P}\{\max_{1\le j,l\le p}\sup_{\tau\in\mathbb{T}}|\frac{1}{n}\sum_{i=1}^{n}X_i^{(j)}X_i^{(l)}1(Q_i<\tau)-E[1(Q_i<\tau)]E[X_i^{(j)}X_i^{(l)}]|$$
$$\ge\left[2\tilde{C}\left(1+\frac{(EM_{XX}^2)^{1/2}\sqrt{\log(p^2n)}}{\sqrt{n}}\right)+1\right]\frac{\sqrt{\log(p^2n)}}{\sqrt{n}}\,|\,(Q_1,\cdots Q_n)\}$$
$$\le\mathbb{P}\left\{\max_{1\le j,l\le p}\max_{1\le k\le n}|\frac{1}{n}\sum_{i=1}^{k}\left(X_i^{(j)}X_i^{(l)}-E[X_i^{(j)}X_i^{(l)}]\right)|\ge\left[2\tilde{C}\left(1+\frac{(EM_{XX}^2)^{1/2}\sqrt{\log(p^2n)}}{\sqrt{n}}\right)+1\right]\frac{\sqrt{\log(p^2n)}}{\sqrt{n}}\right\}$$
$$\le\mathbb{P}\left\{\max_{1\le j,l\le p}\max_{1\le k\le n}|\frac{1}{n}\sum_{i=1}^{n}\left(X_i^{(j)}X_i^{(l)}-E[X_i^{(j)}X_i^{(l)}]\right)\xi_i^{(k)}|\ge\left[2\tilde{C}\left(1+\frac{(EM_{XX}^2)^{1/2}\sqrt{\log(p^2n)}}{\sqrt{n}}\right)+1\right]\frac{\sqrt{\log(p^2n)}}{\sqrt{n}}\right\}$$
$$\le\frac{1}{(p^2n)^{\tilde{C}}}+\tilde{C}\frac{EM_{XX}^2}{(n\log(p^2n))}=o_p(1).$$

By Assumption 1,

$$\left[2\tilde{C}\left(1+\frac{(EM_{XX}^2)^{1/2}\sqrt{\log(p^2n)}}{\sqrt{n}}\right)+1\right]\frac{\sqrt{\log(p^2n)}}{\sqrt{n}}=O_p\left(\frac{\sqrt{\log p}}{\sqrt{n}}+\frac{\sqrt{\log n}}{\sqrt{n}}\right).$$

If I consider $p >> n$,

$$\left[2\tilde{C}\left(1+\frac{(EM_{XX}^2)^{1/2}\sqrt{\log(p^2 n)}}{\sqrt{n}}\right)+1\right]\frac{\sqrt{\log(p^2 n)}}{\sqrt{n}} = O_p\left(\frac{\sqrt{\log p}}{\sqrt{n}}\right).$$

This shows that

$$\sup_{\tau\in\mathbb{T}}\max_{1\le j,l\le p}|\frac{1}{n}\sum_{i=1}^{n}X_i^{(j)}X_i^{(l)}1(Q_i<\tau)-E[(X_i^{(j)}X_i^{(l)}1(Q_i<\tau)]| = O_p(\frac{\sqrt{\log p}}{\sqrt{n}})$$

□

Define $\widehat{\kappa}(s_0,c_0,\mathbb{T},\hat{\Sigma}) = \min_{\tau\in\mathbb{T}}\min_{J_0\subset\{1,\ldots,2p\},|J_0|\le s_0}\min_{\gamma\neq 0,\|\gamma_{J_0^c}\|_1\le c_0\|\gamma_{J_0}\|_1}\frac{(\gamma'\frac{1}{n}\boldsymbol{X}(\tau)'\boldsymbol{X}(\tau)\gamma)^{1/2}}{\|\gamma_{J_0}\|_2}$

and recall Assumption 2 (1.3.3)

$$\kappa(s_0,c_0,\mathbb{T},\Sigma) = \min_{\tau\in\mathbb{T}}\min_{J_0\subset\{1,\ldots,2p\},|J_0|\le s_0}\min_{\gamma\neq 0,\|\gamma_{J_0^c}\|_1\le c_0\|\gamma_{J_0}\|_1}\frac{(\gamma' E[\boldsymbol{X}_i(\tau)'\boldsymbol{X}_i(\tau)]\gamma)^{1/2}}{\|\gamma_{J_0}\|_2} > 0.$$

Define the event

$$\mathbb{A}_5 := \left\{\frac{\kappa(s_0,c_0,\mathbb{T},\hat{\Sigma})^2}{2} < \widehat{\kappa}(c_0,\mathbb{T},\Sigma)^2\right\}$$

The next lemma provides a lower bound on the probability of set $\mathbb{A}_5$.

Lemma 7. *Let Assumptions 1-2 be satisfied,*

$$\mathbb{P}\{\mathbb{A}_5\} \ge 1-(\frac{1}{(p^2)^{\tilde{C}}}+\tilde{C}\frac{EM_{XX}^2}{(n\log p^2)})-(\frac{1}{(p^2 n)^{\tilde{C}}}+\tilde{C}\frac{EM_{XX}^2}{(n\log(p^2 n))}).$$

Proof of Lemma 7. Start with

$$\left|\gamma'\frac{1}{n}\boldsymbol{X}(\tau)'\boldsymbol{X}(\tau)\gamma\right| = \left|\gamma'\left(\frac{1}{n}\boldsymbol{X}(\tau)'\boldsymbol{X}(\tau)-E\left[\boldsymbol{X}_i(\tau)'\boldsymbol{X}_i(\tau)\right]+E\left[\boldsymbol{X}_i(\tau)'\boldsymbol{X}_i(\tau)\right]\right)\gamma\right| \quad \text{(A.2.7)}$$

$$\ge \left|\gamma' E\left[\boldsymbol{X}_i(\tau)'\boldsymbol{X}_i(\tau)\right]\gamma\right| - \left|\gamma'\left(\frac{1}{n}\boldsymbol{X}(\tau)'\boldsymbol{X}(\tau)-E\left[\boldsymbol{X}_i(\tau)'\boldsymbol{X}_i(\tau)\right]\right)\gamma\right| \quad \text{(A.2.8)}$$

by Holders' inequality

$$\left|\gamma'\left(\frac{1}{n}\boldsymbol{X}(\tau)'\boldsymbol{X}(\tau)-E\left[\boldsymbol{X}_i(\tau)'\boldsymbol{X}_i(\tau)\right]\right)\gamma\right| \leq \left\|\gamma\right\|_1^2 \left\|\frac{1}{n}\boldsymbol{X}(\tau)'\boldsymbol{X}(\tau)-E\left[\boldsymbol{X}_i(\tau)'\boldsymbol{X}_i(\tau)\right]\right\|_\infty \tag{A.2.9}$$

Note that I have the restriction set definition

$$\left\|\gamma\right\|_1 \leq \left\|\gamma_{J_0}\right\|_1 + \left\|\gamma_{J_0^c}\right\|_1 \leq (1+c_0)\left\|\gamma_{J_0}\right\|_1 \leq (1+c_0)\sqrt{s_0}\left\|\gamma_{J_0}\right\|_2 \tag{A.2.10}$$

So $\frac{\|\gamma\|_1}{\|\gamma_{J_0}\|_2} \leq (1+c_0)\sqrt{s_0}$. Then divide (A.2.9) by $\left\|\gamma_{J_0}\right\|_2^2$ I have

$$\left|\frac{\left|\gamma'\frac{1}{n}\boldsymbol{X}(\tau)'\boldsymbol{X}(\tau)\gamma\right|}{\left\|\gamma_{J_0}\right\|_2^2} - \frac{\left|\gamma' E\left[\boldsymbol{X}_i(\tau)'\boldsymbol{X}_i(\tau)\right]\gamma\right|}{\left\|\gamma_{J_0}\right\|_2^2}\right| \leq \frac{\left\|\gamma\right\|_1^2}{\left\|\gamma_{J_0}\right\|_2^2}\left\|\frac{1}{n}\boldsymbol{X}(\tau)'\boldsymbol{X}(\tau)-E\left[\boldsymbol{X}_i(\tau)'\boldsymbol{X}_i(\tau)\right]\right\|_\infty \tag{A.2.11}$$

$$\leq (1+c_0)^2 s_0 \left\|\frac{1}{n}\boldsymbol{X}(\tau)'\boldsymbol{X}(\tau)-E\left[\boldsymbol{X}_i(\tau)'\boldsymbol{X}_i(\tau)\right]\right\|_\infty \tag{A.2.12}$$

Since

$$\frac{\left|\gamma' E\left[\boldsymbol{X}_i(\tau)'\boldsymbol{X}_i(\tau)\right]\gamma\right|}{\left\|\gamma_{J_0}\right\|_2^2} - \frac{\left|\gamma'\frac{1}{n}\boldsymbol{X}(\tau)'\boldsymbol{X}(\tau)\gamma\right|}{\left\|\gamma_{J_0}\right\|_2^2} \leq \left|\frac{\left|\gamma'\frac{1}{n}\boldsymbol{X}(\tau)'\boldsymbol{X}(\tau)\gamma\right|}{\left\|\gamma_{J_0}\right\|_2^2} - \frac{\left|\gamma' E\left[\boldsymbol{X}_i(\tau)'\boldsymbol{X}_i(\tau)\right]\gamma\right|}{\left\|\gamma_{J_0}\right\|_2^2}\right| \tag{A.2.13}$$

We obtain

$$\frac{\left|\gamma'\frac{1}{n}\boldsymbol{X}(\tau)'\boldsymbol{X}(\tau)\gamma\right|}{\left\|\gamma_{J_0}\right\|_2^2} \geq \frac{\left|\gamma' E\left[\boldsymbol{X}_i(\tau)'\boldsymbol{X}_i(\tau)\right]\gamma\right|}{\left\|\gamma_{J_0}\right\|_2^2} - (1+c_0)^2 s_0 \left\|\frac{1}{n}\boldsymbol{X}(\tau)'\boldsymbol{X}(\tau)-E\left[\boldsymbol{X}_i(\tau)'\boldsymbol{X}_i(\tau)\right]\right\|_\infty \tag{A.2.14}$$

Minimize over $\tau \in \mathbb{T}$ on right side,

$$\frac{\left|\gamma'\frac{1}{n}\boldsymbol{X}(\tau)'\boldsymbol{X}(\tau)\gamma\right|}{\left\|\gamma_{J_0}\right\|_2^2} \geq \min_{\tau\in\mathbb{T}} \frac{\left|\gamma' E\left[\boldsymbol{X}_i(\tau)'\boldsymbol{X}_i(\tau)\right]\gamma\right|}{\left\|\gamma_{J_0}\right\|_2^2} - (1+c_0)^2 s_0 \sup_{\tau\in\mathbb{T}} \left\|\frac{1}{n}\boldsymbol{X}(\tau)'\boldsymbol{X}(\tau)-E\left[\boldsymbol{X}_i(\tau)'\boldsymbol{X}_i(\tau)\right]\right\|_\infty \tag{A.2.15}$$

Minimize over $\{\gamma \in \mathbb{R}^{2p} \setminus 0\}$ on right side,

$$\text{(A.2.16)} \quad \frac{\left|\gamma' \frac{1}{n} \boldsymbol{X}(\tau)' \boldsymbol{X}(\tau) \gamma\right|}{\left\|\gamma_{J_0}\right\|_2^2} \geq \kappa(c_0, \mathbb{T}, \Sigma)^2 - (1+c_0)^2 s_0 \sup_{\tau \in \mathbb{T}} \left\| \frac{1}{n} \boldsymbol{X}(\tau)' \boldsymbol{X}(\tau) - E\left[\boldsymbol{X}_i(\tau)' \boldsymbol{X}_i(\tau)\right] \right\|_\infty$$

The above inequality is ture for all $\tau \in \mathbb{T}$ and $\{\gamma \in \mathbb{R}^{2p} \setminus 0\}$, so minimize over $\tau \in \mathbb{T}$ and $\{\gamma \in \mathbb{R}^{2p} \setminus 0\}$ on left side I obtain,

$$\text{(A.2.17)} \qquad \hat{\kappa}(c_0, \mathbb{T}, \widehat{\Sigma})^2 \geq \kappa(c_0, \mathbb{T}, \Sigma)^2 - (1+c_0)^2 s_0 \sup_{\tau \in \mathbb{T}} \left\| \frac{1}{n} \boldsymbol{X}(\tau)' \boldsymbol{X}(\tau) - E\left[\boldsymbol{X}_i(\tau)' \boldsymbol{X}_i(\tau)\right] \right\|_\infty$$

So if I can prove that with probability approaching one,

$$(1+c_0)^2 s_0 \sup_{\tau \in \mathbb{T}} \left\| \frac{1}{n} \boldsymbol{X}(\tau)' \boldsymbol{X}(\tau) - E\left[\boldsymbol{X}_i(\tau)' \boldsymbol{X}_i(\tau)\right] \right\|_\infty \leq \frac{\kappa(c_0, \mathbb{T}, \Sigma)^2}{2},$$

that will imply of $\frac{\kappa(c_0, \mathbb{T}, \Sigma)^2}{2} \leq \widehat{\kappa}(c_0, \mathbb{T}, \widehat{\Sigma})^2$ with probability approaching one.

Next, by Lemma 6

$$\sup_{\tau \in \mathbb{T}} \left\| \frac{1}{n} \boldsymbol{X}(\tau)' \boldsymbol{X}(\tau) - E\left[\boldsymbol{X}_i(\tau)' \boldsymbol{X}_i(\tau)\right] \right\|_\infty = O_p\left(\frac{\sqrt{\log p}}{\sqrt{n}}\right)$$

$$\text{(A.2.18)} \quad \mathbb{P}\left\{(1+c_0)^2 s_0 \sup_{\tau \in \mathbb{T}} \left\| \frac{1}{n} \boldsymbol{X}(\tau)' \boldsymbol{X}(\tau) - E\left[\boldsymbol{X}_i(\tau)' \boldsymbol{X}_i(\tau)\right] \right\|_\infty \geq (1+c_0)^2 s_0 \tilde{C} \frac{\sqrt{\log p}}{\sqrt{n}}\right\} = o(1),$$

We get with probability approaching one,

$$(1+c_0)^2 s_0 \sup_{\tau \in \mathbb{T}} \left\| \frac{1}{n} \boldsymbol{X}(\tau)' \boldsymbol{X}(\tau) - E\left[\boldsymbol{X}_i(\tau)' \boldsymbol{X}_i(\tau)\right] \right\|_\infty < (1+c_0)^2 s_0 \tilde{C} \frac{\sqrt{\log p}}{\sqrt{n}} \leq \kappa(c_0, \mathbb{T}, \Sigma)^2/2,$$

since left side of that inequality converges to zero in probability, and the right side is constant. Then by (A.2.18) and (A.2.17)

$$\text{(A.2.19)} \qquad \mathbb{P}\{\mathbb{A}_5\} \geq 1 - o(1).$$

□

Lemma 8. *Suppose that $\delta_0 = 0$. Let Assumption1 amd 2 hold with $\kappa = \kappa(\frac{1+\mu}{1-\mu}, \mathbb{T}, \Sigma)$ for $\mu \in (0,1)$. Let $(\widehat{\alpha}, \widehat{\tau})$ be the LASSO estimator defined by* (1.2.4) *with with $\lambda = \frac{C}{\mu}\frac{\sqrt{\log p}}{\sqrt{n}}$. Then, conditional on events $\mathbb{A}_1$, $\mathbb{A}_2$, $\mathbb{A}_3$, $\mathbb{A}_4$ and $\mathbb{A}_5$, I have*

$$\left\|\widehat{f} - f_0\right\|_n \le \frac{2\sqrt{2}}{\kappa}\left(\sqrt{C_2^2 + \mu\lambda}\right)\sqrt{s_0}\lambda,$$
$$\|\widehat{\alpha} - \alpha_0\|_1 \le \frac{4\sqrt{2}}{(1-\mu)\kappa^2}\frac{C_2^2 + \mu\lambda}{\sqrt{C_3^2 - \mu\lambda}} s_0\lambda.$$

Proof of Lemma 8. Note that $\delta_0 = 0$ implies $\left\|f_{(\alpha_0,\widehat{\tau})} - f_0\right\|^2 = 0$.Conditional on events $\mathbb{A}_1$, $\mathbb{A}_2$, $\mathbb{A}_3$, $\mathbb{A}_4$ with (A.1.20), I have

$$\left\|\widehat{f} - f_0\right\|_n^2 + (1-\mu)\lambda\left\|\widehat{\boldsymbol{D}}(\widehat{\alpha} - \alpha_0)\right\|_1 \le 2\lambda\left\|\widehat{\boldsymbol{D}}(\widehat{\alpha} - \alpha_0)_{J_0}\right\|_1, \tag{A.2.20}$$

which implies that

$$\left\|\widehat{\boldsymbol{D}}(\widehat{\alpha} - \alpha_0)_{J_0^c}\right\|_1 \le \frac{1+\mu}{1-\mu}\left\|\widehat{\boldsymbol{D}}(\widehat{\alpha} - \alpha_0)_{J_0}\right\|_1. \tag{A.2.21}$$

As in Lemma 7, conditional on event $\mathbb{A}_5$, apply Assumption 2, specifically UARE $\kappa = \kappa(\frac{1+\mu}{1-\mu}, \mathbb{T}, \Sigma)$, to yield

$$\begin{aligned}
\kappa^2\left\|\widehat{\boldsymbol{D}}(\widehat{\alpha} - \alpha_0)_{J_0}\right\|_2^2 &\le 2\widehat{\kappa}(\frac{1+\mu}{1-\mu}, \mathbb{T})^2\left\|\widehat{\boldsymbol{D}}(\widehat{\alpha} - \alpha_0)_{J_0}\right\|_2^2 \\
&\le \frac{2}{n}\left\|\boldsymbol{X}(\widehat{\tau})\widehat{\boldsymbol{D}}(\widehat{\alpha} - \alpha_0)\right\|_2^2 \\
&= \frac{2}{n}(\widehat{\alpha} - \alpha_0)'\widehat{\boldsymbol{D}}\boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\widehat{\tau})\widehat{\boldsymbol{D}}(\widehat{\alpha} - \alpha_0) \\
&\le \frac{2\max(\widehat{\boldsymbol{D}})^2}{n}(\widehat{\alpha} - \alpha_0)'\boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\widehat{\tau})(\widehat{\alpha} - \alpha_0) \\
&= 2\max(\widehat{\boldsymbol{D}})^2\left\|\widehat{f} - f_0\right\|_n^2,
\end{aligned} \tag{A.2.22}$$

where the last equality is due to the assumption that $\delta_0 = 0$.

Combining (A.2.20) with (A.2.22) yields

$$\begin{aligned}\left\|\widehat{f}-f_0\right\|_n^2 &\le \left\|\widehat{f}-f_0\right\|_n^2+\left(1-\mu\right)\lambda\left\|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)\right\|_1 \le 2\lambda\left\|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)_{J_0}\right\|_1\\ &\le 2\lambda\sqrt{s_0}\left\|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)_{J_0}\right\|_2 \le \frac{2\sqrt{2}\lambda}{\kappa}\sqrt{s_0}\max(\widehat{\boldsymbol{D}})\left\|\widehat{f}-f_0\right\|_n.\end{aligned}$$

Cancel $\left\|\widehat{f}-f_0\right\|_n$ on the both sides of the inequality,

$$\left\|\widehat{f}-f_0\right\|_n \le \frac{2\sqrt{2}\lambda}{\kappa}\sqrt{s_0}\max(\widehat{\boldsymbol{D}})$$

then conditional on $\mathbb{A}_1$,

$$\left\|\widehat{f}-f_0\right\|_n \le \frac{2\sqrt{2}}{\kappa}\left(\sqrt{C_2^2+\mu\lambda}\right)\sqrt{s_0}\lambda.$$

Next, conditional on $\mathbb{A}_1$, $\mathbb{A}_3$, $\mathbb{A}_4$ and $\mathbb{A}_5$, by (A.2.21)

$$\begin{aligned}\left\|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)\right\|_1 &= \left\|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)_{J_0}\right\|_1+\left\|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)_{J_0^c}\right\|_1\\ &\le 2\left(1-\mu\right)^{-1}\left\|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)_{J_0}\right\|_1\\ &\le 2\left(1-\mu\right)^{-1}\sqrt{s_0}\left\|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)_{J_0}\right\|_2\\ &\le \frac{2}{\kappa\left(1-\mu\right)}\sqrt{s_0}\max(\widehat{\boldsymbol{D}})\left\|\widehat{f}-f_0\right\|_n\\ &\le \frac{4\sqrt{2}\lambda}{\left(1-\mu\right)\kappa^2}s_0(\max(\widehat{\boldsymbol{D}})^2)\\ &\le \frac{4\sqrt{2}\lambda}{\left(1-\mu\right)\kappa^2}s_0(C_2^2+\mu\lambda),\end{aligned} \tag{A.2.23}$$

which proves the second conclusion of the lemma, since conditional on $\mathbb{A}_4$

$$\left\|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)\right\|_1 \ge \min(\widehat{\boldsymbol{D}})\left\|\widehat{\alpha}-\alpha_0\right\|_1 \ge \sqrt{C_3^2-\mu\lambda}\left\|\widehat{\alpha}-\alpha_0\right\|_1. \tag{A.2.24}$$

$$\left\|\widehat{\alpha}-\alpha_0\right\|_1 \le \frac{4\sqrt{2}}{\left(1-\mu\right)\kappa^2}\frac{C_2^2+\mu\lambda}{\sqrt{C_3^2-\mu\lambda}}s_0\lambda. \tag{A.2.25}$$

Hence, the second conclusion of lemma follows give the lower bound on the probability of

$\mathbb{A}_1 \cap \mathbb{A}_2 \cap \mathbb{A}_3 \cap \mathbb{A}_4 \cap \mathbb{A}_5$. □

Proof of Theorem 1. The proof follows immediately from combining Assumption 1 and 2 with Lemma 8. In particular,

$$\begin{aligned}\mathbb{P}\{\mathbb{A}_1 \cap \mathbb{A}_2 \cap \mathbb{A}_3 \cap \mathbb{A}_4 \cap \mathbb{A}_5\} \geq 1 &- \left(\frac{1}{p^{\tilde{C}_1}} + \tilde{C}_2 \frac{EM^2_{X^2}}{n \log p}\right) - \left(\frac{1}{p^{\tilde{C}_3}} + \tilde{C}_4 \frac{EM^2_{X t_0}}{n \log p}\right) \\ &- \left(\frac{1}{p^{\tilde{C}_5}} + \tilde{C}_6 \frac{EM^2_{UX}}{n \log p}\right) - \left(\frac{1}{(pn)^{\tilde{C}_7}} + \tilde{C}_8 \frac{EM^2_{UX}}{n \log (pn)}\right) \\ &- \left(\frac{1}{p^{2\tilde{C}_9}} + \tilde{C}_{10} \frac{EM^2_{XX}}{n \log p^2}\right) - \left(\frac{1}{(p^2 n)^{\tilde{C}_{11}}} + \tilde{C}_{12} \frac{EM^2_{XX}}{n \log (p^2 n)}\right).\end{aligned}$$

□

A.3 Proofs for Section 1.3.2

The following lemma gives an upper bound of $|\widehat{\tau} - \tau_0|$ using only Assumption 3, conditional on the events $\mathbb{A}_1$, $\mathbb{A}_2$, $\mathbb{A}_3$ and $\mathbb{A}_4$.

Lemma 9. *Suppose that Assumption 3 holds. Let*

$$\eta^* = \max\left\{\min_i |Q_i - \tau_0|, \frac{1}{C_4}\left(2C_1(3+\mu)\left(C_2^2 + \mu\lambda\right)^{\frac{1}{2}} s_0 \lambda\right)\right\}$$

where C_4 is the constant defined in Assumption 3. Then conditional on the events $\mathbb{A}_1$, $\mathbb{A}_2$, $\mathbb{A}_3$ and $\mathbb{A}_4$

$$|\widehat{\tau} - \tau_0| \leq \eta^*.$$

Proof of Lemma 9. As in the proof of Lemma 5, I have, on the events $\mathbb{A}_1$,$\mathbb{A}_2$,$\mathbb{A}_3$ and $\mathbb{A}_4$

$$\begin{aligned}&\widehat{S}_n - S_n(\alpha_0, \tau_0) \\ &= \left\|\widehat{f} - f_0\right\|_n^2 - 2n^{-1}\sum_{i=1}^n U_i X_i'(\widehat{\beta} - \beta_0) - 2n^{-1}\sum_{i=1}^n U_i X_i'(\widehat{\delta} - \delta_0) 1(Q_i < \widehat{\tau}) - R_n \qquad \text{(A.3.1)} \\ &\geq \left\|\widehat{f} - f_0\right\|_n^2 - \mu\lambda \left\|\widehat{\boldsymbol{D}}(\widehat{\alpha} - \alpha_0)\right\|_1 - R_n.\end{aligned}$$

Then $\mathbb{A}_1$, $\mathbb{A}_2$, $\mathbb{A}_3$ and $\mathbb{A}_4$,

$$
\begin{aligned}
&\left[\widehat{S}_n+\lambda\left\|\widehat{\boldsymbol{D}}\widehat{\alpha}\right\|_1\right]-\left[S_n(\alpha_0,\tau_0)+\lambda\|\boldsymbol{D}\alpha_0\|_1\right]\\
&\geq\left\|\widehat{f}-f_0\right\|_n^2-\lambda\left\|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)\right\|_1-\lambda\left[\|\boldsymbol{D}\alpha_0\|_1-\left\|\widehat{\boldsymbol{D}}\widehat{\alpha}\right\|_1\right]-R_n\\
&\quad\text{using (A.1.25) and (A.1.26)}\\
&\geq\left\|\widehat{f}-f_0\right\|_n^2-2\lambda\left\|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)_{J_0}\right\|_1-\lambda\left|\|\boldsymbol{D}\alpha_0\|_1-\left\|\widehat{\boldsymbol{D}}\widehat{\alpha}\right\|_1\right|-R_n\\
&\quad\text{using (A.1.28), (A.1.29) and (A.1.30) to bound the last three terms,}\\
&\geq\left\|\widehat{f}-f_0\right\|_n^2-\left(6\lambda\sqrt{C_2^2+\mu\lambda}C_1s_0+2\mu\lambda\sqrt{C_2^2+\mu\lambda}C_1s_0\right)\\
&\geq\left\|\widehat{f}-f_0\right\|_n^2-\left(2C_1(3+\mu)\left(C_2^2+\mu\lambda\right)^{\frac{1}{2}}s_0\lambda\right)\geq 0\\
&\quad\text{by Lemma 1.}
\end{aligned}
\tag{A.3.2}
$$

Suppose now that $|\hat{\tau}-\tau_0|>\eta^*$, then Assumption 3 and (A.3.2) together imply that

$$
\left[\widehat{S}_n+\lambda\left\|\widehat{\boldsymbol{D}}\widehat{\alpha}\right\|_1\right]-\left[S_n(\alpha_0,\tau_0)+\lambda\|\boldsymbol{D}\alpha_0\|_1\right]\geq\left\|\widehat{f}-f_0\right\|_n^2-C_4\eta^*>0,
$$

which leads to contradiction as $\widehat{\tau}$ is the minimizer of (1.2.4). Therefore, I have proved the lemma. □

The following lemma demonstrates that if my design satisfies Assumption 1, Assumption 4 is automatic in my case.

Lemma 10 (Assumption 4)**.** *If Assumption 1 is satisfied, then for any $\eta>C\frac{\log p}{n}>0$, with $C>0$, there exists a finite constant $C_5<\infty$, such that*

$$
\mathbb{P}\left\{\sup_{1\leq j,l\leq p}\sup_{|\tau-\tau_0|<\eta}\frac{1}{n}\sum_{i=1}^{n}\left|X_i^{(j)}X_i^{(l)}\right||1(Q_i<\tau_0)-1(Q_i<\tau)|\leq C_5\eta\right\}\to 1 \tag{A.3.3}
$$

$$
\mathbb{P}\left\{\sup_{|\tau-\tau_0|<\eta}\left|\frac{1}{n}\sum_{i=1}^{n}U_iX_i'\delta_0[1(Q_i<\tau_0)-1(Q_i<\tau)]\right|\leq\frac{\lambda\sqrt{\eta}}{2}\right\}\to 1, \tag{A.3.4}
$$

as $n\to\infty$.

Proof of Lemma 10. Recall the matrix $\Xi_{n,n}$, as defined in (A.1.15), and let $\xi_i^{(k)}$ represent the element in the $i-th$ row and $k-th$ column of this matrix. Additionally, define $\tilde{\xi}_i^{(q)}$ to be the element in the $i-th$ row and $q-th$ column of the transpose of $\Xi_{n,n}$, denoted as $\Xi_{n,n}^T$.

Sort $\{X_i, U_i, Q_i\}_{i=1}^n$ by $(Q_1, \cdots Q_n)$ in ascending order, then

$$\begin{aligned}
&\mathbb{P}\left\{\sup_{1\le j,l\le p}\sup_{|\tau-\tau_0|<\eta}\frac{1}{n}\sum_{i=1}^{n}\left|X_i^{(j)}X_i^{(l)}\right||1(Q_i<\tau_0)-1(Q_i<\tau)|>C_5\eta\right\}\\
=&\mathbb{P}\left\{\max_{1\le j,l\le p}\max_{[n(\tau_0-\eta)]\le q\le k\le[n(\tau_0+\eta)]}\frac{1}{n}\sum_{i=q}^{k}\left|X_i^{(j)}X_i^{(l)}\xi_i^{(k)}\xi_i^{(q)}\right|>C_5\eta\right\}
\end{aligned}$$

then $\left|X_i^{(j)}X_i^{(l)}\xi_i^{(k)}\xi_i^{(q)}\right|$ is independent random variable(not identical) and

$$\max_{1\le j,l\le p}\max_{[n(\tau_0-\eta)]\le q\le i\le k\le[n(\tau_0+\eta)]}\left|X_i^{(j)}X_i^{(l)}\xi_i^{(k)}\xi_i^{(q)}\right|\le\max_{1\le j,l\le p}\max_{1\le i\le n}\left|X_i^{(j)}X_i^{(l)}\right|\le EM_{XX}+C_2^2<\infty.$$

So under Assumption 1, conditions for maximal inequalities are stratified. Lemma E.4 (ii)of Chernozhukov et al. (2017) provides (set $\eta=1$ and $s=2$ in their Lemma):

(A.3.5)

$$\begin{aligned}
&\mathbb{P}\left\{\max_{1\le j,l\le p}\max_{[n(\tau_0-\eta)]\le q\le k\le[n(\tau_0+\eta)]}\sum_{i=q}^{k}\left|X_i^{(j)}X_i^{(l)}\xi_i^{(k)}\xi_i^{(q)}\right|\ge 2E\left[\max_{1\le j,l\le p}\max_{[n(\tau_0-\eta)]\le q\le k\le[n(\tau_0+\eta)]}\sum_{i=q}^{k}\left|X_i^{(j)}X_i^{(l)}\xi_i^{(k)}\xi_i^{(q)}\right|\right]+t\right\}\\
\le&\tilde{C}\frac{E\left[\max_{1\le j,l\le p}\max_{[n(\tau_0-\eta)]\le q\le i\le k\le[n(\tau_0+\eta)]}\left|X_i^{(j)}X_i^{(l)}\xi_i^{(k)}\xi_i^{(q)}\right|\right]}{t^2},\text{ set } t=\log p\text{, then I have}\\
\le&\tilde{C}\frac{(EM_{XX}+C_2^2)^2}{(\log p)^2}
\end{aligned}$$

Lemma E.3 of Chernozhukov et al. (2017) provides

$$\begin{aligned}
&E\left[\max_{1\le j,l\le p}\max_{[n(\tau_0-\eta)]\le q\le k\le[n(\tau_0+\eta)]}\sum_{i=q}^{k}\left|X_i^{(j)}X_i^{(l)}\xi_i^{(k)}\xi_i^{(q)}\right|\right]\\
\le&\tilde{C}\max_{1\le j,l\le p}\max_{[n(\tau_0-\eta)]\le q\le k\le[n(\tau_0+\eta)]}E\left[\sum_{i=q}^{k}\left|X_i^{(j)}X_i^{(l)}\xi_i^{(k)}\xi_i^{(q)}\right|\right]\qquad\text{(A.3.6)}\\
+&\tilde{C}E\left[\max_{1\le j,l\le p}\max_{[n(\tau_0-\eta)]\le q\le i\le k\le[n(\tau_0+\eta)]}\left|X_i^{(j)}X_i^{(l)}\xi_i^{(k)}\xi_i^{(q)}\right|\right]log(2n\eta\cdot p^2)\\
\le&\tilde{C}(EM_{XX}+C_2^2)[(2n\eta)+log(2n\eta\cdot p^2)]
\end{aligned}$$

Combining (A.3.6) with (A.3.5),

$$\mathbb{P}\left\{\max_{1\le j,l\le p}\max_{[n(\tau_0-\eta)]\le q\le k\le[n(\tau_0+\eta)]}\frac{1}{n}\sum_{i=q}^{k}\left|X_i^{(j)}X_i^{(l)}\xi_i^{(k)}\tilde{\xi}_i^{(q)}\right|\ge 2\tilde{C}(EM_{XX}+C_2^2)[(2\eta)+\frac{log(2n\eta\cdot p^2)}{n}]+\frac{\log p}{n}\right\}$$
$$\le\tilde{C}\frac{(EM_{XX}+C_2^2)^2}{(\log p)^2}.$$

There exist a positive constant C_5 such that $2\tilde{C}(EM_{XX}+C_2^2)[(2\eta)+\frac{log(2n\eta\cdot p^2)}{n}]+\frac{\log p}{n}=C_5\eta$ given $\eta\ge C\frac{\log p}{n}$, then

$$\mathbb{P}\left\{\sup_{1\le j,l\le p}\sup_{|\tau-\tau_0|<\eta}\frac{1}{n}\sum_{i=1}^{n}\left|X_i^{(j)}X_i^{(l)}\right||1(Q_i<\tau_0)-1(Q_i<\tau)|>C_5\eta\right\}\le\tilde{C}\frac{(EM_{XX}+C_2^2)^2}{(\log p)^2}=o_p(1).$$

Hence, I have proved (A.3.3); (A.3.4) can be proven using parallelly arguments.

Sort $\{X_i,U_i,Q_i\}_{i=1}^n$ by $(Q_1,\cdots Q_n)$ in ascending order, then

$$\mathbb{P}\left\{\sup_{|\tau-\tau_0|<\eta}\left|\frac{1}{n}\sum_{i=1}^{n}U_iX_i'\delta_0[1(Q_i<\tau_0)-1(Q_i<\tau)]\right|\ge\frac{\lambda\sqrt{\eta}}{2}\right\}$$
$$=\mathbb{P}\left\{\max_{[n(\tau_0-\eta)]\le q\le k\le[n(\tau_0+\eta)]}\left|\frac{1}{n}\sum_{i=q}^{k}U_iX_i'\delta_0\xi_i^{(k)}\tilde{\xi}_i^{(l)}\right|\ge\frac{\lambda\sqrt{\eta}}{2}\right\}$$

then $U_iX_i'\delta_0\xi_i^{(k)}\tilde{\xi}_i^{(l)}$ is independent centered random variable(not identical),

$$\max_{[n(\tau_0-\eta)]\le q\le k\le[n(\tau_0+\eta)]}\sum_{i=q}^{k}Var[U_iX_i'\delta_0\xi_i^{(k)}\tilde{\xi}_i^{(l)}]\le 2n\eta Var[U_iX_i\delta_0]\le 2n\eta\|\delta_0\|_1^2EM_{UX}^2$$

and

$$\max_{[n(\tau_0-\eta)]\le q\le k\le[n(\tau_0+\eta)]}|U_iX_i'\delta_0\xi_i^{(k)}\tilde{\xi}_i^{(l)}|\le\|\delta_0\|_1\max_{1\le i\le n}\max_{1\le j\le p}|U_iX_i^{(j)}|<\|\delta_0\|_1M_{UX}<\infty.$$

So under Assumption 1, conditions for maximal inequalities are stratified. Lemma E.2(ii) of Chernozhukov et al. (2017) provides (set $\eta=1$ and $s=2$ in their Lemma):

(A.3.7)

$$\mathbb{P}\left\{\max_{[n(\tau_0-\eta)]\le q\le k\le[n(\tau_0+\eta)]}|\sum_{i=q}^{k}U_iX_i'\delta_0\xi_i^{(k)}\tilde{\xi}_i^{(l)}|\ge 2E\left[\max_{[n(\tau_0-\eta)]\le q\le k\le[n(\tau_0+\eta)]}|\sum_{i=q}^{k}U_iX_i'\delta_0\xi_i^{(k)}\tilde{\xi}_i^{(l)}|\right]+t\right\}$$

$$\le\exp\{-\frac{t^2}{3\max_{[n(\tau_0-\eta)]\le q\le k\le[n(\tau_0+\eta)]}|U_iX_i'\delta_0\xi_i^{(k)}\tilde{\xi}_i^{(l)}|}\}+\tilde{C}\frac{\|\delta_0\|_1^2EM_{UX}^2}{t^2},\ \text{set } t=(n\eta\log p)^{1/2}\text{,then I have}$$

$$\le\frac{1}{(p)^{\tilde{C}}}+\tilde{C}\frac{\|\delta_0\|_1^2EM_{UX}^2}{(n\eta\log p)}$$

Lemma E.1 of Chernozhukov et al. (2017) provides

(A.3.8)

$$E\left[\max_{[n(\tau_0-\eta)]\le q\le k\le[n(\tau_0+\eta)]}|\sum_{i=q}^{k}U_iX_i'\delta_0\xi_i^{(k)}\tilde{\xi}_i^{(l)}|\right]$$

$$\le\tilde{C}\sqrt{\max_{[n(\tau_0-\eta)]\le q\le k\le[n(\tau_0+\eta)]}\sum_{i=q}^{k}Var[U_iX_i'\delta_0\xi_i^{(k)}\tilde{\xi}_i^{(l)}]}\sqrt{\log(2n\eta)^2}+\tilde{C}\|\delta_0\|_1\sqrt{EM_{UX}^2}\log(2n\eta)^2$$

$$\le\tilde{C}\|\delta_0\|_1\sqrt{2n\eta EM_{UX}^2}\sqrt{\log(2n\eta)^2}+\tilde{C}\|\delta_0\|_1\sqrt{EM_{UX}^2}\log(2n\eta)^2$$

$$=\tilde{C}\|\delta_0\|_1\sqrt{EM_{UX}^2}[\sqrt{2n\eta\log(2n\eta)^2}+\log(2n\eta)^2]$$

Combining (A.3.8) with (A.3.7),

(A.3.9)

$$\mathbb{P}\left\{\max_{[n(\tau_0-\eta)]\le q\le k\le[n(\tau_0+\eta)]}|\frac{1}{n}\sum_{i=q}^{k}U_iX_i'\delta_0\xi_i^{(k)}\tilde{\xi}_i^{(l)}|\ge 2\tilde{C}\|\delta_0\|_1\sqrt{EM_{UX}^2}[\frac{\sqrt{2\eta\log(2n\eta)^2}}{\sqrt{n}}+\frac{\log(2n\eta)^2}{n}]+\frac{\sqrt{\eta\log p}}{\sqrt{n}}\right\}$$

$$\le\frac{1}{(p)^{\tilde{C}}}+\tilde{C}\frac{\|\delta_0\|_1^2EM_{UX}^2}{(n\eta\log p)}.$$

Set λ by (1.3.1),there exist a positive constant C such that $2\tilde{C}\|\delta_0\|_1\sqrt{EM_{UX}^2}[\frac{\sqrt{2\eta\log(2n\eta)^2}}{\sqrt{n}}+\frac{\log(2n\eta)^2}{n}]+\frac{\sqrt{\eta\log p}}{\sqrt{n}}=\frac{C}{2\mu}\frac{\sqrt{\log p}}{\sqrt{n}}\sqrt{\eta}$ given $\eta\ge C\frac{\log p}{n}$,

$$\mathbb{P}\left\{\sup_{|\tau-\tau_0|<\eta}\left|\frac{1}{n}\sum_{i=1}^{n}U_iX_i'\delta_0[1(Q_i<\tau_0)-1(Q_i<\tau)]\right|\ge\frac{C}{2\mu}\frac{\sqrt{\log p}}{\sqrt{n}}\sqrt{\eta}\right\}$$

$$\le\frac{1}{(p)^{\tilde{C}}}+\tilde{C}\frac{\|\delta_0\|_1^2EM_{UX}^2}{(n\eta\log p)}=o_p(1)$$

Hence, I have proved (A.3.4). □

We now provide a lemma for bounding the prediction loss as well as the l_1 estimation loss for α_0.To do so, I define a constant G_2, and functions of $(\lambda, c_\alpha, c_\tau, \|\delta_0\|_1)$ G_1 and G_3 accordingly:

$$
\begin{aligned}
G_2 &= \frac{12\left(C_2^2+\mu\lambda\right)}{\kappa^2},\\
G_1 &= \sqrt{c_\tau} + \left(2\sqrt{C_3^2-\mu\lambda}\right)^{-1} C_5 \|\delta_0\|_1 c_\tau,\\
G_3 &= \frac{2\sqrt{2}\left(C_2^2+\mu\lambda\right)^{\frac{1}{2}}\sqrt{C_5 C_1}}{\kappa}(c_\alpha c_\tau)^{1/2}.
\end{aligned}
$$

Lemma 11. *Suppose that* $|\hat{\tau}-\tau_0| \le c_\tau$ *and* $\|\hat{\alpha}-\alpha_0\|_1 \le c_\alpha$ *for some* (c_τ, c_α). *Let Assumption 2 and 4 hold with* $\mathbb{S} = \{|\tau-\tau_0| \le c_\tau\}$, $\kappa = \kappa(\frac{2+\mu}{1-\mu}, \mathbb{S}, \Sigma)$ *for* $0<\mu<1$. *Then, conditional on* $\mathbb{A}_1$,$\mathbb{A}_2$, $\mathbb{A}_3$, $\mathbb{A}_4$ *and* $\mathbb{A}_5$ *, I have*

$$
\begin{aligned}
\left\|\widehat{f}-f_0\right\|_n^2 &\le 3\lambda\cdot\left\{G_1 \vee G_2\lambda s_0 \vee G_3\sqrt{s_0\|\delta_0\|_1}\right\},\\
\|\widehat{\alpha}-\alpha_0\|_1 &\le \frac{3}{(1-\mu)\sqrt{C_3^2-\mu\lambda}}\cdot\left\{G_1 \vee G_2\lambda s_0 \vee G_3\sqrt{s_0\|\delta_0\|_1}\right\}.
\end{aligned}
$$

Proof of Lemma 11. Note that

$$
|R_n| = \left|2n^{-1}\sum_{i=1}^{n} U_i X_i'\delta_0\left\{1(Q_i<\widehat{\tau})-1(Q_i<\tau_0)\right\}\right| \le \lambda\sqrt{c_\tau}. \tag{A.3.10}
$$

by Assumption 4 (1.3.7). Conditioning on $\mathbb{A}_4$, the triangular inequality implies that

$$
\begin{aligned}
\left|\left\|\widehat{\boldsymbol{D}}\alpha_0\right\|_1 - \|\boldsymbol{D}\alpha_0\|_1\right| &\leq \left|\sum_{j=1}^{p}\left(\left\|X^{(j)}(\hat{\tau})\right\|_n - \left\|X^{(j)}(\tau_0)\right\|_n\right)\left|\delta_0^{(j)}\right|\right| \\
&\quad \text{applying the mean value theorem to } \left\|X^{(j)}(\hat{\tau})\right\|_n \\
&\leq \sum_{j=1}^{p}\left(2\left\|X^{(j)}(t_0)\right\|_n\right)^{-1}\left|\left\|X^{(j)}(\hat{\tau})\right\|_n^2 - \left\|X^{(j)}(\tau_0)\right\|_n^2\right|\left|\delta_0^{(j)}\right| \\
&\leq \sum_{j=1}^{p}\left(2\left\|X^{(j)}(t_0)\right\|_n\right)^{-1}\left|\delta_0^{(j)}\right|\frac{1}{n}\sum_{i=1}^{n}\left|X_i^{(j)}\right|^2\left|\mathbf{1}\{Q_i<\hat{\tau}\}-\mathbf{1}\{Q_i<\tau_0\}\right| \\
&\leq \left(2\sqrt{C_3^2-\mu\lambda}\right)^{-1}\|\delta_0\|_1 C_5 c_\tau.
\end{aligned}
\tag{A.3.11}
$$

where the last inequality is due to Assumption 4(1.3.5). We now consider two cases:
(i) $\left\|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)_{J_0}\right\|_1 > \sqrt{c_\tau} + \left(2\sqrt{C_3^2-\mu\lambda}\right)^{-1} C_5\|\delta_0\|_1 c_\tau$ and
(ii) $\left\|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)_{J_0}\right\|_1 \leq \sqrt{c_\tau} + \left(2\sqrt{C_3^2-\mu\lambda}\right)^{-1} C_5\|\delta_0\|_1 c_\tau$.

Case (i): Combine (A.3.10) and (A.3.11)

$$
\lambda\left|\left\|\widehat{\boldsymbol{D}}\alpha_0\right\|_1 - \|\boldsymbol{D}\alpha_0\|_1\right| + R_n < \lambda\left(2\sqrt{C_3^2-\mu\lambda}\right)^{-1}\|\delta_0\|_1 C_5(c_\tau+\lambda\sqrt{c_\tau}) + \lambda\sqrt{c_\tau} < \lambda\left\|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)_{J_0}\right\|_1.
$$

Combine the above results with (A.1.27), I have

$$
\left\|\widehat{f}-f_0\right\|_n^2 + (1-\mu)\lambda\left\|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)\right\|_1 \leq 3\lambda\left\|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)_{J_0}\right\|_1, \tag{A.3.12}
$$

which implies

$$
(1-\mu)\left\|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)\right\|_1 \leq 3\left\|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)_{J_0}\right\|_1.
$$

Then subtract $(1-\mu)\left\|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)_{J_0}\right\|_1$ on both sides,

$$
\left\|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)_{J_0^c}\right\|_1 \leq \frac{2+\mu}{1-\mu}\left\|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)_{J_0}\right\|_1. \tag{A.3.13}
$$

With the adaptive restricted eigenvalue condition $\kappa(s_0, \frac{2+\mu}{1-\mu}, \mathbb{S}, \Sigma)$, I apply Assumption 2 in this case.

Recall that $\boldsymbol{X}_i(\tau) = (X_i', X_i'\mathbf{1}\{Q_i < \tau\})'$ and $\boldsymbol{X}(\tau) = (\boldsymbol{X}_1(\tau)', \cdots, \boldsymbol{X}_n(\tau)')'$.Note the fact that

$$\frac{1}{n}\sum_{i=1}^{n}\left\{2\Big(\boldsymbol{X}_i(\widehat{\tau})'\widehat{\alpha} - \boldsymbol{X}_i(\widehat{\tau})'\alpha_0\Big)\Big(X_i'\delta_0\big[1(Q_i<\tau_0) - 1(Q_i<\widehat{\tau})\big]\Big)\right\}$$
$$=2\Big(\widehat{\alpha}'\boldsymbol{X}(\widehat{\tau})' - \alpha_0'\boldsymbol{X}(\widehat{\tau})'\Big)\Big(X'\delta_0\big[1(Q_i<\tau_0) - 1(Q_i<\widehat{\tau})\big]\Big)$$
$$=2\Big(\widehat{\alpha}'\boldsymbol{X}_i(\widehat{\tau})' - \alpha_0'\boldsymbol{X}_i(\widehat{\tau})'\Big)\Big(\boldsymbol{X}_i(\tau_0)\alpha_0 - \boldsymbol{X}_i(\widehat{\tau})\alpha_0\Big)$$
$$=2\widehat{\alpha}'\boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\tau_0)\alpha_0 - 2\alpha_0'\boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\tau_0)\alpha_0 - 2\widehat{\alpha}'\boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\widehat{\tau})\alpha_0 + 2\alpha_0'\boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\widehat{\tau})\alpha_0$$
$$=-2\alpha_0'\boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\tau_0)\alpha_0 + 2\alpha_0'\boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\widehat{\tau})\alpha_0 + 2\alpha_0'\boldsymbol{X}(\tau_0)'\boldsymbol{X}(\widehat{\tau})\widehat{\alpha} - 2\alpha_0'\boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\widehat{\tau})\widehat{\alpha}$$
$$\text{since } \alpha_0'\boldsymbol{X}(\tau_0)'\boldsymbol{X}(\tau_0)\alpha_0 + \alpha_0'\boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\widehat{\tau})\alpha_0 \geq 2\alpha_0'\boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\tau_0)\alpha_0$$
$$\geq -\alpha_0'\boldsymbol{X}(\tau_0)'\boldsymbol{X}(\tau_0)\alpha_0 - \alpha_0'\boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\widehat{\tau})\alpha_0 + 2\alpha_0'\boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\widehat{\tau})\alpha_0 + 2\alpha_0'\boldsymbol{X}(\tau_0)'\boldsymbol{X}(\widehat{\tau})\widehat{\alpha} - 2\alpha_0'\boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\widehat{\tau})\widehat{\alpha}$$
$$=-\alpha_0'\boldsymbol{X}(\tau_0)'\boldsymbol{X}(\tau_0)\alpha_0 + \alpha_0'\boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\widehat{\tau})\alpha_0 + 2\alpha_0'\boldsymbol{X}(\tau_0)'\boldsymbol{X}(\widehat{\tau})\widehat{\alpha} - 2\alpha_0'\boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\widehat{\tau})\widehat{\alpha}$$

Since it is assumed that $|\widehat{\tau} - \tau_0| \leq c_\tau$, Assumption 2 only needs to hold with $\mathbb{S}$ in the c_τ neighborhood of τ_0. As $\delta_0 \neq 0$, (A.2.22) now has an extra term

$$\kappa^2 \left\|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)_{J_0}\right\|_2^2 \leq 2\widehat{\kappa}(\frac{2+\mu}{1-\mu}, \mathbb{S}, \widehat{\Sigma})^2 \left\|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)_{J_0}\right\|_2^2$$
$$\leq \frac{2}{n}\left\|\boldsymbol{X}(\widehat{\tau})\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)\right\|_2^2$$
$$= \frac{2}{n}(\widehat{\alpha}-\alpha_0)'\widehat{\boldsymbol{D}}\boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\widehat{\tau})\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)$$
$$\leq \frac{2\left\|\widehat{\boldsymbol{D}}\right\|_\infty^2}{n}(\widehat{\alpha}-\alpha_0)'\boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\widehat{\tau})(\widehat{\alpha}-\alpha_0)$$
$$\leq 2\left\|\hat{f}-f_0\right\|_n^2\left\|\widehat{\boldsymbol{D}}\right\|_\infty^2\left(\left\|\hat{f}-f_0\right\|_n^2 - \alpha_0'\boldsymbol{X}(\tau_0)'\boldsymbol{X}(\tau_0)\alpha_0 + \alpha_0'\boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\widehat{\tau})\alpha_0 + 2\alpha_0'\boldsymbol{X}(\tau_0)'\boldsymbol{X}(\widehat{\tau})\widehat{\alpha} - 2\alpha_0'\boldsymbol{X}(\widehat{\tau})'\boldsymbol{X}(\widehat{\tau})\widehat{\alpha}\right)$$
$$\leq 2\left\|\widehat{\boldsymbol{D}}\right\|_\infty^2\left\|\hat{f}-f_0\right\|_n^2 + 2\left\|\widehat{\boldsymbol{D}}\right\|_\infty^2\frac{1}{n}\sum_{i=1}^{n}\left\{2\Big(\boldsymbol{X}_i(\widehat{\tau})'\widehat{\alpha} - \boldsymbol{X}_i(\widehat{\tau})'\alpha_0\Big)\Big(X_i'\delta_0\big[1(Q_i<\tau_0) - 1(Q_i<\widehat{\tau})\big]\Big)\right\}$$
$$\leq 2\left\|\widehat{\boldsymbol{D}}\right\|_\infty^2\left(\left\|\hat{f}-f_0\right\|_n^2 + 2c_\alpha\|\delta_0\|_1\sup_j\frac{1}{n}\sum_{i=1}^{n}\left|X_i^{(j)}\right|^2|1(Q_i<\tau_0) - 1(Q_i<\widehat{\tau})|\right)$$
$$\leq 2\left(C_2^2+\mu\lambda\right)\left(\left\|\widehat{f}-f_0\right\|_n^2 + 2C_5\|\delta_0\|_1 c_\alpha c_\tau\right),$$

where the last inequality is due to events $\mathbb{A}_1$ and Assumption 4(1.3.5). Combining this result with (A.3.12), I have

$$\begin{aligned}\left\|\widehat{f}-f_0\right\|_n^2 &\leq 3\lambda\left\|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)_{J_0}\right\|_1\\ &\leq 3\lambda\sqrt{s_0}\left\|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)_{J_0}\right\|_2\\ &\leq 3\lambda\sqrt{s_0}\left(2\kappa^{-2}\left(C_2^2+\mu\lambda\right)\left(\left\|\widehat{f}-f_0\right\|_n^2+2C_5\|\delta_0\|_1 c_\alpha c_\tau\right)\right)^{1/2}.\end{aligned}$$

Applying $a+b\leq 2a\vee 2b$, I get the upper bound of $\left\|\widehat{f}-f_0\right\|_n$ on $\mathbb{A}_1$,$\mathbb{A}_2$, $\mathbb{A}_3$, $\mathbb{A}_4$, $\mathbb{A}_5$, as

$$\left\|\widehat{f}-f_0\right\|_n^2 \leq \frac{36\left(C_2^2+\mu\lambda\right)}{\kappa^2}\lambda^2 s_0 \vee \frac{6\sqrt{2}\left(C_2^2+\mu\lambda\right)^{\frac{1}{2}}\sqrt{C_5 C_1}}{\kappa}\lambda\sqrt{s_0\|\delta_0\|_1}(c_\alpha c_\tau)^{1/2}. \tag{A.3.14}$$

To derive the upper bound for $\|\widehat{\alpha}-\alpha_0\|_1$, use (A.3.13) ,

$$\begin{aligned}\min(\widehat{\boldsymbol{D}})\|\widehat{\alpha}-\alpha_0\|_1 \leq \left\|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)\right\|_1 &\leq \frac{3}{1-\mu}\left\|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)_{J_0}\right\|_1\\ &\leq \frac{3}{1-\mu}\sqrt{s_0}\left\|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)_{J_0}\right\|_2\\ &\leq \frac{3}{1-\mu}\sqrt{s_0}\left(2\kappa^{-2}\left(C_2^2+\mu\lambda\right)\left(\left\|\widehat{f}-f_0\right\|_n^2+2c_\alpha c_\tau C_5\|\delta_0\|_1\right)\right)^{1/2}\\ &= \frac{3\sqrt{2}}{(1-\mu)\kappa}\sqrt{s_0}\left(\left(C_2^2+\mu\lambda\right)\left(\left\|\widehat{f}-f_0\right\|_n^2+2C_5\|\delta_0\|_1 c_\alpha c_\tau\right)\right)^{1/2}.\end{aligned}$$

where the last inequality is due to conditional on $\mathbb{A}_3$. Then using the inequality that $a+b\leq 2a\vee 2b$ with (A.2.25) and (A.3.14) yields

$$\|\widehat{\alpha}-\alpha_0\|_1 \leq \frac{36}{(1-\mu)\kappa^2}\frac{\left(C_2^2+\mu\lambda\right)}{\sqrt{C_3^2-\mu\lambda}}\lambda s_0 \vee \frac{6\sqrt{2}}{(1-\mu)\kappa}\frac{\sqrt{C_2^2+\mu\lambda}\sqrt{C_5}}{\sqrt{C_3^2-\mu\lambda}}\sqrt{s_0\|\delta_0\|_1}(c_\alpha c_\tau)^{1/2}.$$

Case (ii): In this case, (A.1.27) shows

$$\left\|\widehat{f}-f_0\right\|_n^2+(1-\mu)\lambda\left\|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)_{J_0^c}\right\|_1 \leq 2\lambda\left\|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)_{J_0}\right\|_1-(1-\mu)\lambda\left\|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)_{J_0}\right\|_1+\lambda\left|\left\|\widehat{\boldsymbol{D}}\alpha_0\right\|_1-\|\boldsymbol{D}\alpha_0\|_1\right|+R_n \tag{A.3.15}$$

which implies

$$\left\|\widehat{f}-f_0\right\|_n^2+\leq(1+\mu)\lambda\left\|\widehat{\boldsymbol{D}}(\widehat{\alpha}-\alpha_0)_{J_0}\right\|_1+\lambda\left|\left\|\widehat{\boldsymbol{D}}\alpha_0\right\|_1-\|\boldsymbol{D}\alpha_0\|_1\right|+R_n. \tag{A.3.16}$$

$$\left\|\widehat{f}-f_0\right\|_n^2 \leq 3\lambda\left(\sqrt{c_\tau}+\left(2\sqrt{C_3^2-\mu\lambda}\right)^{-1} C_5\|\delta_0\|_1 c_\tau\right),$$
$$\|\widehat{\alpha}-\alpha_0\|_1 \leq \frac{3}{(1-\mu)\sqrt{C_3^2-\mu\lambda}}\left(\sqrt{c_\tau}+\left(2\sqrt{C_3^2-\mu\lambda}\right)^{-1} C_5\|\delta_0\|_1 c_\tau\right),$$

which provides the result. □

The following lemma shows that the bound for $|\hat{\tau}-\tau_0|$ can be further tightened if I combine results obtained in Lemmas 9 and 11.

Lemma 12. *Suppose that* $|\hat{\tau}-\tau_0| \leq c_\tau$ *and* $\|\widehat{\alpha}-\alpha_0\|_1 \leq c_\alpha$ *for some* (c_τ, c_α). *Let*

$$\tilde{\eta} = C_4^{-1}\lambda\left(\left(1+\mu\right)\sqrt{C_2^2+\mu\lambda}\,c_\alpha + G_1\right).$$

If Assumption 3 holds, then conditional on the events $\mathbb{A}_1$, $\mathbb{A}_2$, $\mathbb{A}_3$, $\mathbb{A}_4$, *I have,*

$$|\hat{\tau}-\tau_0| \leq \tilde{\eta}.$$

Proof of Lemma 12. Note that on $\mathbb{A}_1$, $\mathbb{A}_2$, $\mathbb{A}_3$ $\mathbb{A}_4$ and Assumption 4(1.3.7),

$$\begin{aligned} &\left|\frac{2}{n}\sum_{i=1}^{n}\left[U_i X_i'\left(\hat{\beta}-\beta_0\right)+U_i X_i' 1\left(Q_i<\hat{\tau}\right)\left(\hat{\delta}-\delta_0\right)\right]\right| \\ \leq\ & \mu\lambda\left(\sqrt{C_2^2+\mu\lambda}\right)\|\hat{\alpha}-\alpha_0\|_1 \leq \mu\lambda\left(\sqrt{C_2^2+\mu\lambda}\right)c_\alpha \end{aligned}$$

and

$$\left|\frac{2}{n}\sum_{i=1}^{n} U_i X_i' \delta_0\left[1\left(Q_i<\hat{\tau}\right)-1\left(Q_i<\tau_0\right)\right]\right| \leq \lambda\sqrt{c_\tau}.$$

Suppose $\tilde{\eta} < |\hat{\tau}-\tau_0| \leq c_\tau$. As in (A.3.1),

$$\widehat{S}_n - S_n(\alpha_0,\tau_0) \geq \left\|\widehat{f}-f_0\right\|_n^2 - \mu\lambda\left(\sqrt{C_2^2+\mu\lambda}\,c_\alpha\right) - \lambda\sqrt{c_\tau}.$$

Furthermore, I obtain

$$\begin{aligned}
&\left[\widehat{S}_n + \lambda \left\|\widehat{\boldsymbol{D}}\widehat{\alpha}\right\|_1\right] - \left[S_n(\alpha_0, \tau_0) + \lambda \|\boldsymbol{D}\alpha_0\|_1\right] \\
&\text{using triangle inequality on } \left\|\widehat{\boldsymbol{D}}\widehat{\alpha}\right\|_1 - \|\boldsymbol{D}\alpha_0\|_1 \\
\geq &\left\|\widehat{f} - f_0\right\|_n^2 - \mu\lambda\left(\sqrt{C_2^2 + \mu\lambda}\, c_\alpha\right) - 2\|\delta_0\|_1 \lambda \sqrt{c_\tau} - \lambda\left(\left\|\widehat{\boldsymbol{D}}(\widehat{\alpha} - \alpha_0)\right\|_1 + \left\|(\widehat{\boldsymbol{D}} - \boldsymbol{D})\alpha_0\right\|_1\right) \\
> &C_4\tilde{\eta} - \left(\left(1+\mu\right)\left(\sqrt{C_2^2 + \mu\lambda}\, c_\alpha\right) + G_1\right)\lambda,
\end{aligned}$$

where the last inequality is due to Assumption 3, Hölder's inequality and (A.3.11).

Since $C_4\tilde{\eta} = \left(\left(1+\mu\right)\sqrt{C_2^2 + \mu\lambda}\, c_\alpha + G_1\right)\lambda$ by definitation, similarly as in the proof of Lemma 9, proof by contradiction yields the result. □

Lemma 11 provides us with three different bounds for $\|\alpha - \alpha_0\|_1$ and the two terms G_1 amd G_3 are functions of c_τ and c_α. If I can show that the bound for $|\hat{\tau} - \tau_0|$ and $|\hat{\alpha} - \alpha_0|$ in 11 and 12 are further tightened, it is useful to apply Lemmas 11 and 12 iteratively. to tighten up the bounds i

Lemma 9 results in that I can start the iteration with $c_\tau^{(0)} = \frac{2C_1(3+\mu)\left(C_2^2+\mu\lambda\right)^{\frac{1}{2}}}{C_4} s_0\lambda$. (1.3.2) in Lemma 1 allow us to choose $c_\alpha^{(0)} = \frac{\left(2C_1(3+\mu)\right)\left(C_2^2+\mu\lambda\right)^{\frac{1}{2}}}{(1-\mu)\left(C_3^2-\mu\lambda\right)^{\frac{1}{2}}} s_0$.

Lemma 13. *Suppose that Assumption1 to 4 hold with* $\mathbb{S} = \left\{|\tau - \tau_0| \leq \eta^*\right\}$, $\kappa = \kappa(s_0, \frac{2+\mu}{1-\mu}, \mathbb{S}, \Sigma)$ *for* $0 < \mu < 1$. *Let* $(\widehat{\alpha}, \widehat{\tau})$ *be the LASSO estimator defined by* (1.2.4) *with* λ *given by* (1.3.1). *In addition, there exists a sequence of constants* $\eta_1, ..., \eta_{m^*}$ *for some finite* m^*. *With probability at least* $1 - \left(\frac{1}{p^{\tilde{C}_1}} + \tilde{C}_2 \frac{EM_{X2}^2}{n\log p}\right) - \left(\frac{1}{p^{\tilde{C}_3}} + \tilde{C}_4 \frac{EM_{Xt_0}^2}{n\log p}\right) - \left(\frac{1}{p^{\tilde{C}_5}} + \tilde{C}_6 \frac{EM_{UX}^2}{n\log p}\right) - \left(\frac{1}{(pn)^{\tilde{C}_7}} + \tilde{C}_8 \frac{EM_{UX}^2}{(n\log pn)}\right) - \left(\frac{1}{(p^2)^{\tilde{C}_9}} + \tilde{C}_{10} \frac{EM_{XX}^2}{(n\log p^2)}\right) - \left(\frac{1}{(p^2 n)^{\tilde{C}_{11}}} + \tilde{C}_{12} \frac{EM_{XX}^2}{(n\log p^2 n)}\right)$ *I have*

$$\begin{aligned}
\left\|\widehat{f} - f_0\right\|_n^2 &\leq 3G_2\lambda^2 s_0, \\
\|\widehat{\alpha} - \alpha_0\|_1 &\leq \frac{3}{(1-\mu)\sqrt{C_3^2 - \mu\lambda}} G_2\lambda s_0, \\
|\hat{\tau} - \tau_0| &\leq \left(\frac{3\left(1+\mu\right)\sqrt{\left(C_2^2 + \mu\lambda\right)}}{(1-\mu)\sqrt{\left(C_3^2 - \mu\lambda\right)}} + 1\right)\frac{1}{C_4} G_2\lambda^2 s_0.
\end{aligned}$$

Proof of Lemma 13. The iteration to implement is as follows:

Step 1: Starting values $c_\tau^{(0)} = \frac{2C_1(3+\mu)\left(C_2^2+\mu\lambda\right)^{\frac{1}{2}}}{C_4} s_0\lambda$ and $c_\alpha^{(0)} = \frac{\left(2C_1(3+\mu)\right)\left(C_2^2+\mu\lambda\right)^{\frac{1}{2}}}{\left(1-\mu\right)\left(C_3^2-\mu\lambda\right)^{\frac{1}{2}}} s_0$.

Step 2:When $m \geq 1$,

$$\begin{aligned}
G_1^{(m-1)} =& \sqrt{c_\tau^{(m-1)}} + \left(2\sqrt{C_3^2-\mu\lambda}\right)^{-1} C_5\|\delta_0\|_1 c_\tau^{(m-1)},\\
G_3^{(m-1)} =& \frac{2\sqrt{2}\left(C_2^2+\mu\lambda\right)^{\frac{1}{2}}\sqrt{C_5 C_1}}{\kappa}\sqrt{c_\alpha^{(m-1)} c_\tau^{(m-1)}},\\
c_\alpha^{(m)} =& \frac{3}{(1-\mu)\sqrt{C_3^2-\mu\lambda}}\cdot\left\{G_1^{(m-1)} \vee G_2\lambda s_0 \vee G_3^{(m-1)}\sqrt{s_0\|\delta_0\|_1}\right\},\\
c_\tau^{(m)} =& \frac{\lambda}{C_4}\left(\left(1+\mu\right)\sqrt{C_2^2+\mu\lambda}\, c_\alpha^{(m)} + G_1^{(m-1)}\right).
\end{aligned}$$

Step 3: We stop the iteration if

$$\left\{G_1^{(m)} \vee G_2\lambda s_0 \vee G_3^{(m)}\sqrt{s_0\|\delta_0\|_1}\right\}$$

doesn't change.

Suppose step 3 met under $\left\{G_1^{(m)} \vee G_2\lambda s_0 \vee G_3^{(m)}\sqrt{s_0\|\delta_0\|_1}\right\} = G_2\lambda s_0$, then the bound in the lemma is reached within m^*, a finite number, of iterative applications.

Since $G_1^{(m-1)}$ and $G_2\lambda s_0$ are positive, $\frac{G_1^{(m-1)}}{G_2\lambda s_0} > 0$. Note that $c_\alpha^{(m)} \geq \frac{3}{(1-\mu)\sqrt{C_3^2-\mu\lambda}} G_2\lambda s_0$, I have

$$\begin{aligned}
c_\tau^{(m)} =& \frac{\lambda}{C_4}\left(\left(1+\mu\right)\sqrt{C_2^2+\mu\lambda}\, c_\alpha^{(m)} + G_1^{(m-1)}\right)\\
\geq& \frac{\lambda}{C_4}\left(\frac{3\left(1+\mu\right)\sqrt{C_2^2+\mu\lambda}}{(1-\mu)\sqrt{C_3^2-\mu\lambda}} G_2\lambda s_0 + G_1^{(m-1)}\right)\\
\geq& \frac{1}{C_4}\left(\frac{3\left(1+\mu\right)\sqrt{C_2^2+\mu\lambda}}{(1-\mu)\sqrt{C_3^2-\mu\lambda}} + \frac{G_1^{(m-1)}}{G_2\lambda s_0}\right) G_2\lambda^2 s_0\\
>& \frac{1}{C_4}\left(\frac{3\left(1+\mu\right)\sqrt{C_2^2+\mu\lambda}}{(1-\mu)\sqrt{C_3^2-\mu\lambda}}\right) G_2\lambda^2 s_0
\end{aligned} \tag{A.3.17}$$

Note that (A.3.17) shows that $c_\tau^{(m)} \geq C s_0 \frac{\log p}{2n}$ are valid for all each application of Lemma 9 to Lemma 12. Then $c_\alpha^{(m^*+1)}$ is the bound given in the statement of the lemma for $\|\hat{\alpha}-\alpha_0\|_1$. Next,

$$
\begin{aligned}
c_\tau^{(m^*+1)} =& \frac{\lambda}{C_4}\left(\left(1+\mu\right)\sqrt{C_2^2+\mu\lambda}\,c_\alpha^{(m^*+1)} + G_1^{(m^*)}\right) \\
\leq& \frac{\lambda}{C_4}\left(\frac{3\left(1+\mu\right)\sqrt{C_2^2+\mu\lambda}}{(1-\mu)\sqrt{C_3^2-\mu\lambda}} G_2\lambda s_0 + G_2\lambda s_0\right) \\
=& \left(\frac{3\left(1+\mu\right)\sqrt{\left(C_2^2+\mu\lambda\right)}}{(1-\mu)\sqrt{\left(C_3^2-\mu\lambda\right)}}+1\right)\frac{G_2}{C_4}\lambda^2 s_0,
\end{aligned}
$$

which is the bound given in the statement of the lemma for $|\widehat{\tau}-\tau_0|$.

Next, I turn to proof-of-existence for m^*. First, by induction I can show that $G_1^{(m-1)}$, $G_1^{(m-1)}$, $c_\alpha^{(m)}$ and $c_\tau^{(m)}$ are decreasing as m increases. We start the iteration with setting of $c_\tau^{(0)}$ and $c_\alpha^{(0)}$ in step 1. By step 2, as long as $n, p,\ s_0$ and $\|\delta_0\|_1$ are large enough, I obtain (in the following derivation, $\tilde{C}$ are different constant in each term,but all positive and finite)

$$
\begin{aligned}
G_1^{(0)} =& \sqrt{c_\tau^{(0)}} + \left(2\sqrt{C_3^2-\mu\lambda}\right)^{-1} C_5\|\delta_0\|_1 c_\tau^{(0)} = \tilde{C}\sqrt{s_0\lambda} + \tilde{C}\|\delta_0\|_1 s_0\lambda, \\
G_3^{(0)} =& \frac{2\sqrt{2}\left(C_2^2+\mu\lambda\right)^{\frac{1}{2}}\sqrt{C_5C_1}}{\kappa}\sqrt{c_\alpha^{(0)}}\sqrt{c_\tau^{(0)}} = \tilde{C}\sqrt{s_0^2\lambda},
\end{aligned}
$$

$$
\text{Then } \left\{G_1^{(0)} \vee G_2\lambda s_0 \vee G_3^{(0)}\sqrt{s_0\|\delta_0\|_1}\right\} = G_3^{(0)}\sqrt{s_0\|\delta_0\|_1},
$$

follows from $\|\delta_0\|_1 s_0\lambda = o_p(1)$.

$$
\begin{aligned}
c_\alpha^{(1)} =& \frac{3}{(1-\mu)\sqrt{C_3^2-\mu\lambda}}\cdot\left\{G_1^{(0)} \vee G_2\lambda s_0 \vee G_3^{(0)}\sqrt{s_0\|\delta_0\|_1}\right\} = \tilde{C}s_0\sqrt{s_0\|\delta_0\|_1\lambda}, \\
c_\tau^{(1)} =& \frac{\lambda}{C_4}\left(\left(1+\mu\right)\sqrt{C_2^2+\mu\lambda}\,c_\alpha^{(1)} + G_1^{(0)}\right) = \tilde{C}s_0\lambda\sqrt{s_0\|\delta_0\|_1\lambda} + \tilde{C}\lambda\sqrt{s_0\lambda} + \tilde{C}\|\delta_0\|_1 s_0\lambda^2.
\end{aligned}
$$

Thus I have

$$
c_\alpha^{(0)} > c_\alpha^{(1)} \text{ and } c_\tau^{(0)} > c_\tau^{(1)}.
$$

We assume

$$
c_\alpha^{(m)} > c_\alpha^{(m+1)} \text{ and } c_\tau^{(m)} > c_\tau^{(m+1)},
$$

it is easy to show

$$G_1^{(m)} > G_1^{(m+1)} \text{ and } G_3^{(m)} > G_3^{(m+1)}$$

then

$$c_\alpha^{(m+1)} > c_\alpha^{(m+2)} \text{ and } c_\tau^{(m+1)} > c_\tau^{(m+2)}.$$

This means that applying the iteration can tighten up the bounds.

We use proof by contradiction to be shown that there exist m^* such that

$$\left\{G_1^{(m^*)} \vee G_2\lambda s_0 \vee G_3^{(m^*)}\sqrt{s_0\|\delta_0\|_1}\right\} = G_2\lambda s_0.$$

Suppose for all $m > 1$,

$$\left\{G_1^{(m)} \vee G_3^{(m)}\sqrt{s_0\|\delta_0\|_1}\right\} > G_2\lambda s_0.$$

As $G_1^{(m-1)}$, $G_3^{(m-1)}$ are decreasing as m increases, and $\left\{G_1^{(m)} \vee G_3^{(m)}\sqrt{s_0\|\delta_0\|_1}\right\}$ is bounded, there are two cases to consider:

Case (1):

$$G_1^{(m)} \leq G_3^{(m)}\sqrt{s_0\|\delta_0\|_1}$$

for m sufficiently large. Let $G_3^{(m)}$ converge to $G_3^{(\infty)}$ and $G_3^{(\infty)} > G_2\lambda s_0$.

$$c_\alpha^{(\infty)} = \frac{3}{(1-\mu)\sqrt{C_3^2-\mu\lambda}} G_3\sqrt{s_0\|\delta_0\|_1} =: H_1\sqrt{s_0\|\delta_0\|_1}\sqrt{c_\alpha^{(\infty)}}\sqrt{c_\tau^{(\infty)}},$$

where H_1 is defined accordingly as

$$H_1 = \frac{6\sqrt{2}\left(C_2^2+\mu\lambda\right)^{\frac{1}{2}}\sqrt{C_5C_1}}{(1-\mu)\sqrt{C_3^2-\mu\lambda}\kappa}.$$

$$c_\alpha^\infty = H_1^2 s_0\|\delta_0\|_1 c_\tau^\infty,$$

$$\begin{aligned}
c_\tau^\infty =& C_4^{-1}\lambda\left(\left(1+\mu\right)\sqrt{\left(C_2^2+\mu\lambda\right)}c_\alpha^\infty + \sqrt{c_\tau^\infty} + \left(2\sqrt{C_3^2-\mu\lambda}\right)^{-1} C_5\|\delta_0\|_1 c_\tau^\infty\right)\\
=& C_4^{-1}\left(1+\mu\right)\sqrt{\left(C_2^2+\mu\lambda\right)}\lambda c_\alpha^\infty + C_4^{-1}\lambda\sqrt{c_\tau^\infty} + C_4^{-1}\left(2\sqrt{C_3^2-\mu\lambda}\right)^{-1} C_5\|\delta_0\|_1\lambda c_\tau^\infty\\
=&: H_2\lambda c_\alpha^\infty + H_3\lambda\sqrt{c_\tau^\infty} + H_4\|\delta_0\|_1\lambda c_\tau^\infty,
\end{aligned}$$

by defining

$$H_2 =: C_4^{-1}\left(1+\mu\right)\sqrt{\left(C_2^2+\mu\lambda\right)},$$

$$H_3 =: C_4^{-1},$$

$$H_4 =: C_4^{-1}\left(2\sqrt{C_3^2-\mu\lambda}\right)^{-1} C_5.$$

To solve the above equation system, as n, p are sufficiently large, $\sqrt{C_3^2-\mu\lambda}$ and $\sqrt{C_2^2+\mu\lambda}$ converge to constants; $s_0\|\delta\|_1\lambda$ and $\|\delta_0\|_1\lambda$ converge to 0,

$$\begin{aligned}
c_\tau^\infty =& \left(\frac{H_1^2H_2s_0\|\delta\|_1\lambda^2+H_3\lambda}{1-H_1^2H_2s_0\|\delta\|_1\lambda-H_4\lambda\|\delta\|_1}\right)^2 = O_p(\lambda^2),\\
c_\alpha^{(\infty)} =& H_1^2 s_0\|\delta_0\|_1 c_\tau^\infty = O_p(s_0\|\delta_0\|_1\lambda^2).
\end{aligned}$$

Then,

$$G_3^{(\infty)}\sqrt{s_0\|\delta_0\|_1} = \frac{(1-\mu)\sqrt{C_3^2-\mu\lambda}}{3} c_\alpha^{(\infty)} = O_p(s_0\|\delta_0\|_1\lambda^2),$$

Obviously, the above leads to contradiction, because $c_\tau^\infty < s_0\lambda^2$ and $G_3^{(\infty)}\sqrt{s_0\|\delta_0\|_1} < G_2\lambda s_0$.

Case (2):

$$G_1^{(m)} > G_3^{(m)}\sqrt{s_0\|\delta_0\|_1}$$

for m sufficiently large. Let $G_1^{(m)}$ converge to $G_1^{(\infty)}$ and $G_1^{(\infty)} > G_2\lambda s_0$.

Thus, I have that

$$
\begin{aligned}
c_{\alpha}^{(\infty)} &= G_1 \frac{3}{(1-\mu)\sqrt{C_3^2-\mu\lambda}}, \\
c_{\tau}^{(\infty)} &= C_4^{-1}\lambda\Big(\big(1+\mu\big)\sqrt{\big(C_2^2+\mu\lambda\big)}c_{\alpha}^{(\infty)}+G_1^{(\infty)}\Big) \\
&= C_4^{-1}\lambda\left(\big(1+\mu\big)\sqrt{\big(C_2^2+\mu\lambda\big)}\frac{3}{(1-\mu)\sqrt{C_3^2-\mu\lambda}}+1\right)G_1^{(\infty)} \\
&= C_4^{-1}\left(\frac{3\big(1+\mu\big)\sqrt{\big(C_2^2+\mu\lambda\big)}}{\big(1-\mu\big)\sqrt{C_3^2-\mu\lambda}}+1\right)\lambda\sqrt{c_{\tau}^{(\infty)}} \\
&+ C_4^{-1}\left(\frac{3\big(1+\mu\big)\sqrt{\big(C_2^2+\mu\lambda\big)}}{\big(1-\mu\big)\sqrt{C_3^2-\mu\lambda}}+1\right)\Big(2\sqrt{C_3^2-\mu\lambda}\Big)^{-1}C_5\|\delta_0\|_1\lambda c_{\tau}^{(\infty)} \\
&=: H_5\lambda\sqrt{c_{\tau}^{(\infty)}}+H_6\|\delta_0\|_1\lambda c_{\tau}^{(\infty)},
\end{aligned}
$$

where H_5 and H_6 are defined accordingly. Furthermore, as n, p are sufficiently large, $\sqrt{C_3^2-\mu\lambda}$ and $\sqrt{C_2^2+\mu\lambda}$ converge to constants , $\|\delta_0\|_1\lambda$ converges to 0,

$$
c_{\tau}^{\infty} \quad = \quad \left(\frac{H_5\lambda}{1-H_6\|\delta_0\|_1\lambda}\right)^2 = O_p(\lambda^2).
$$

Then

$$
G_1^{(\infty)} = \left(1+\Big(2\sqrt{C_3^2-\mu\lambda}\Big)^{-1}\lambda\|\delta_0\|_1 C_5\right)\sqrt{c_{\tau}^{(\infty)}}+\Big(2\sqrt{C_3^2-\mu\lambda}\Big)^{-1}C_5\|\delta_0\|_1 c_{\tau}^{(\infty)} = O_p(\lambda+\lambda^2),
$$

which leads to contradiction, because $c_{\tau}^{\infty} < s_0\lambda^2$ and $G_1^{(\infty)} < G_2\lambda s_0$.

Finally, Lemma 11 yields

$$
\left\|\widehat{f}-f_0\right\|_n^2 \leq 3G_2\lambda^2 s_0.
$$

□

Proof of Theorem 2. The proof follows immediately from combining Assumption 1 to 4 with

Lemma 13. In particular,

$$\mathbb{P}\{\mathbb{A}_1 \cap \mathbb{A}_2 \cap \mathbb{A}_3 \cap \mathbb{A}_4 \cap \mathbb{A}_5\} \geq 1 - \left(\frac{1}{p^{\tilde{C}_1}} + \tilde{C}_2 \frac{EM^2_{X^2}}{n \log p}\right) - \left(\frac{1}{p^{\tilde{C}_3}} + \tilde{C}_4 \frac{EM^2_{X t_0}}{n \log p}\right)$$
$$- \left(\frac{1}{p^{\tilde{C}_5}} + \tilde{C}_6 \frac{EM^2_{UX}}{n \log p}\right) - \left(\frac{1}{(pn)^{\tilde{C}_7}} + \tilde{C}_8 \frac{EM^2_{UX}}{n \log(pn)}\right)$$
$$- \left(\frac{1}{p^{2\tilde{C}_9}} + \tilde{C}_{10} \frac{EM^2_{XX}}{n \log p^2}\right) - \left(\frac{1}{(p^2 n)^{\tilde{C}_{11}}} + \tilde{C}_{12} \frac{EM^2_{XX}}{n \log(p^2 n)}\right).$$

□

A.4 Proof of Asymptotic Properties of Nodewise Regression Estimator

The proof is similar to Lemma A.9 in the Appendix of Caner and Kock (2018). We adapt their proof to my purpose.

Define

$$\mathbb{A}_{node} = \{\max_{j+p \in H} \sup_{\tau \in \mathbb{T}} \|X^{(-j)}(\tau)' \upsilon^{(j)}/n\|_\infty \leq \frac{\mu \lambda_{node}}{2}\},$$

$$\mathbb{A}^{(j)}_{EV} = \{\frac{\kappa(s_j, c_0, \mathbb{T}, \hat{M}_{-j,-j})^2}{2} \leq \hat{\kappa}(s_j, c_0, \mathbb{T}, M_{-j,-j})^2\}.$$

$$\mathbb{B}_{node} = \{\max_{j \in H \text{or} j+p \in H} \sup_{\tau \in \mathbb{T}} \|\tilde{X}^{(-j)}(\tau)' \tilde{\upsilon}^{(j)}/n\|_\infty \leq \frac{\mu \lambda_{node}}{2}\},$$

$$\mathbb{B}^{(j)}_{EV} = \{\frac{\kappa(s_j, c_0, \mathbb{T}, \hat{N}_{-j,-j})^2}{2} \leq \hat{\kappa}(s_j, c_0, \mathbb{T}, N_{-j,-j})^2\}.$$

The above four series of events are uniformly on $\tau \in \mathbb{T}$.

Lemma 14. *Let Assumptions 1-5 be satisfied and set* $\lambda_{node} = \frac{C}{\mu}\sqrt{\frac{\log p}{n}}$. *Suppose that* $\widehat{\delta}(\widehat{\tau}) \neq 0$ *estimated via* (1.2.4). *Then*

$$\mathbb{P}\left\{\mathbb{A}_{node} \cap_{j+p \in H} \mathbb{A}^{(j)}_{EV} \cap \mathbb{B}_{node} \cap_{j \in H or j+p \in H} \mathbb{B}^{(j)}_{EV}\right\} \geq 1 - o_p(1).$$

Proof of Lemma 14. To prove probability of event $\mathbb{A}^C_{node}$, I adapt the the proof of Lemma 4 to my purpose,

$$\mathbb{P}\left\{\mathbb{A}_{node}^{C}\right\} = \mathbb{P}\left\{\max_{j+p\in H}\sup_{\tau\in\mathbb{T}}\|X^{(-j)}(\tau)'\upsilon^{(j)}/n\|_{\infty} \le \frac{\mu\lambda_{node}}{2}\right\}$$
$$=\mathbb{P}\left\{\max_{j+p\in H}\max_{1\le l\le p-1}\sup_{\tau\in\mathbb{T}}\frac{1}{n}\sum_{i=1}^{n}X_i^{(-j,l)}(\tau)\upsilon_i^{(j)} \le \frac{\mu\lambda_{node}}{2}\right\}$$

Sort $\{X_i, U_i, Q_i\}_{i=1}^{n}$ by $(Q_1, \cdots Q_n)$ in ascending order, then

$$\mathbb{P}\left\{\max_{j+p\in H}\sup_{\tau\in\mathbb{T}}\max_{1\le l\le p-1}\frac{1}{n}\sum_{i=1}^{n}X_i^{(-j,l)}(\tau)\upsilon_i^{(j)} \le \frac{\mu\lambda_{node}}{2}\right\}$$
$$=\mathbb{P}\left\{\max_{j+p\in H}\max_{1\le k\le n}\max_{1\le l\le p-1}\frac{1}{n}\sum_{i=1}^{k}X_i^{(-j,l)}\upsilon_i^{(j)} \le \frac{\mu\lambda_{node}}{2}\right\}.$$

Given that there are three layers k, j, l across $\max_{1\le k\le n}\max_{j+p\in H}\max_{1\le l\le p-1}\sum_{i=1}^{k}X_i^{(-j,l)}\upsilon_i^{(j)}$, we combine equations (A.1.16) and (A.1.17) while setting $t = \sqrt{n\log((p-1)hn)}$,

$$\mathbb{P}\{\max_{j+p\in H}\max_{1\le k\le n}\max_{1\le l\le p-1}\frac{1}{n}\sum_{i=1}^{k}X_i^{(-j,l)}\upsilon_i^{(j)} \ge$$
$$2\tilde{C}[\frac{\sqrt{n\log((p-1)hn)}}{n} + \frac{\sqrt{EM_{X\upsilon}^2}\log((p-1)hn)}{n}] + \frac{\sqrt{n\log((p-1)hn)}}{n}\}$$
$$\le\frac{1}{[(p-1)hn]^{\tilde{C}}} + \tilde{C}\frac{EM_{X\upsilon}^2}{n\log((p-1)hn)} = o_p(1).$$

We see that

$$2\tilde{C}[\frac{\sqrt{n\log((p-1)hn)}}{n} + \frac{\sqrt{EM_{X\upsilon}^2}\log((p-1)hn)}{n}] + \frac{\sqrt{n\log((p-1)hn)}}{n}$$
$$\le(2\tilde{C}+1)\frac{\sqrt{n\log p^3}}{n} + 2\tilde{C}\frac{\sqrt{EM_{X\upsilon}^2}\log p^3}{n}$$
$$\le\sqrt{\frac{\log p}{n}}((2\tilde{C}+1)\sqrt{3} + 6\tilde{C}\sqrt{\frac{EM_{X\upsilon}^2\log p}{n}})$$

provided that I can find some constant $\tilde{C} > 0$.

Therefore if I choose $\frac{\mu\lambda_{node}}{2} = \sqrt{\frac{\log p}{n}}((2\tilde{C}+1)\sqrt{3} + 6\tilde{C}\sqrt{\frac{EM_{X\upsilon}^2\log p}{n}})$, the same rate as (1.3.1),

$$\mathbb{P}\left\{\mathbb{A}_{node}^{C}\right\} \le \frac{1}{[(p-1)hn]^{\tilde{C}}} + \tilde{C}\frac{EM_{X\upsilon}^2}{n\log((p-1)hn)} = o_p(1)$$

Using analogous arguments to that discussed above with regard to the transpose of $\Xi_{n,n}$ in (A.1.15) and $\tilde{\xi} i^{(l)}$ as the element in the i-th row and $l-th$ column of the transpose of $\Xi n, n$, I can conclude the following inequality:

$$\mathbb{P}\left\{\mathbb{B}^{C}_{node}\right\} \leq \frac{1}{[(p-1)hn]^{\tilde{C}}} + \tilde{C}\frac{EM^2_{Xv}}{n\log((p-1)hn)} = o_p(1)$$

Next, I bound the probability of event $(\cap_{j+p\in H}\mathbb{A}^{(j)}_{EV})^C$ and $(\cap_{j\in H \text{or} j+p\in H}\mathbb{B}^{(j)}_{EV})^C$. Note the fact that for each $j+p\in H$

$(1+c_0)^2 s_j \sup_{\tau\in\mathbb{T}} \|\widehat{M}_{-j,-j}(\tau) - M_{-j,-j}(\tau)\|_\infty \leq (1+c_0)^2 \bar{s} \sup_{\tau\in\mathbb{T}} \|\widehat{M}(\tau) - M(\tau)\|_\infty \leq \frac{\kappa(\bar{s}, c_0, \mathbb{T}, M)}{2} \leq \frac{\kappa(s_j, c_0, \mathbb{T}, M)}{2}$ implies that

$$\left\{(1+c_0)^2 s_j \sup_{\tau\in\mathbb{T}} \|\widehat{M}_{-j,-j}(\tau) - M_{-j,-j}(\tau)\|_\infty \leq \frac{\kappa(s_j, c_0, \mathbb{T}, M)}{2}\right\} \subset \mathbb{A}^{(j)}_{EV}.$$

Thus,

$$\left\{(1+c_0)^2 \bar{s} \sup_{\tau\in\mathbb{T}} \|\widehat{M}(\tau) - M(\tau)\|_\infty \leq \frac{\kappa(\bar{s}, c_0, \mathbb{T}, M)}{2}\right\} \subset \cap_{j+p\in H}\mathbb{A}^{(j)}_{EV}.$$

Then by arguments exactly parallel to those in Lemma 7, I can show,

$$\mathbb{P}\left\{(\cap_{j+p\in H}\mathbb{A}^{(j)}_{EV})^C\right\} \leq o_p(1)$$

provided that $\kappa(s_j, c_0, \mathbb{T}, M) > 0$. Similarly, I can show

$$\mathbb{P}\left\{(\cap_{j\in H \text{or} j+p\in H}\mathbb{B}^{(j)}_{EV})^C\right\} \leq o_p(1)$$

Therefore

$$\mathbb{P}\left\{\mathbb{A}_{node} \cap_{j+p\in H} \mathbb{A}^{(j)}_{EV} \cap \mathbb{B}_{node} \cap_{j\in H \text{or} j+p\in H} \mathbb{B}^{(j)}_{EV}\right\} \geq 1 - o_p(1).$$

□

Proof of Lemma 2. Given $\forall\tau\in\mathbb{T}$ and each $j\in H$ or $j+p\in H$, (1.4.12) is a loss function for linear model, the pointwise oracle inequalities from Theorem 2.4 in van de Geer et al. (2014) for linear model have been proved.

As the uniform oracle inequalities only involve noise conditions $\mathbb{A}_{node}$ and $\mathbb{B}_{node}$, and adaptive restricted eigenvalue conditions $\cap_{j+p\in H}\mathbb{A}^{(j)}_{EV}$ and $\cap_{j\in H \text{or} j+p\in H}\mathbb{B}^{(j)}_{EV}$. Therefore, by Lemma 14, I obtain the following results uniformly in $\mathbb{T}$ and H,

$$\sup_{\tau\in\mathbb{T}}\max_{j+p\in H}\|X^{(-j)}(\tau)'\gamma_j(\tau)-X^{(-j)}(\tau)'\widehat{\gamma}_j(\tau)\|_n \leq \frac{C}{\widehat{\kappa}(\bar{s},c_0,\mathbb{T},\Sigma)}\sqrt{\bar{s}}\lambda_{node} \tag{A.4.1}$$

$$\sup_{\tau\in\mathbb{T}}\max_{j+p\in H}\left\|\gamma_j(\tau)-\widehat{\gamma}_j(\tau)\right\|_1 \leq \frac{C}{\widehat{\kappa}(\bar{s},c_0,\mathbb{T},\Sigma)^2}\bar{s}\lambda_{node} \tag{A.4.2}$$

with probability

$$\mathbb{P}\left\{\mathbb{A}_{node}\cap_{j+p\in H}\mathbb{A}^{(j)}_{EV}\cap\mathbb{B}_{node}\cap_{j\in H \text{or} j+p\in H}\mathbb{B}^{(j)}_{EV}\right\}\geq 1-o(1)$$

In line with the inequalities presented in Lemma A.9 in the Appendix of Caner and Kock (2018), I can establish the following set of inequalities:

$$\max_{j+p\in H}\sup_{\tau\in\mathbb{T}}\|\hat{A}_j(\tau)-A_j(\tau)\|_1 = O_p\left(\bar{s}\sqrt{\frac{\log p}{n}}\right) \tag{A.4.3}$$

$$\max_{j+p\in H}\sup_{\tau\in\mathbb{T}}\|\hat{A}_j(\tau)-A_j(\tau)\|_2 = O_p\left(\sqrt{\frac{\bar{s}\log p}{n}}\right) \tag{A.4.4}$$

$$\max_{j+p\in H}\sup_{\tau\in\mathbb{T}}\|\hat{A}_j(\tau)\|_1 = O_p\left(\sqrt{\bar{s}}\right) \tag{A.4.5}$$

$$\max_{j+p\in H}\sup_{\tau\in\mathbb{T}}\frac{1}{\hat{z}_j(\tau)^2} = O_p(1) \tag{A.4.6}$$

$$\max_{j\in H \text{or} j+p\in H}\sup_{\tau\in\mathbb{T}}\|\hat{B}_j(\tau)-B_j(\tau)\|_1 = O_p\left(\bar{s}\sqrt{\frac{\log p}{n}}\right) \tag{A.4.7}$$

$$\max_{j\in H \text{or} j+p\in H}\sup_{\tau\in\mathbb{T}}\|\hat{B}_j(\tau)-B_j(\tau)\|_2 = O_p\left(\sqrt{\frac{\bar{s}\log p}{n}}\right) \tag{A.4.8}$$

$$\max_{j\in H \text{or} j+p\in H}\sup_{\tau\in\mathbb{T}}\|\hat{B}_j(\tau)\|_1 = O_p\left(\sqrt{\bar{s}}\right) \tag{A.4.9}$$

$$\max_{j\in H \text{or} j+p\in H}\sup_{\tau\in\mathbb{T}}\frac{1}{\widehat{\tilde{z}}_j(\tau)^2} = O_p(1) \tag{A.4.10}$$

We now turn to (1.3.4) and (1.4.20),

$$\max_{j\in H}\sup_{\tau\in\mathbb{T}}\|\hat{\Theta}(\tau)_j - \Theta(\tau)_j\|_1 \leq \max_{j\in H \text{ or } j+p\in H}\sup_{\tau\in\mathbb{T}}\max\{2\|\hat{B}_j(\tau)-B_j(\tau)\|_1, \|\hat{B}_j(\tau)-B_j(\tau)\|_1+\|\hat{A}_j(\tau)-A_j(\tau)\|_1\},$$

$$\max_{j\in H}\sup_{\tau\in\mathbb{T}}\|\hat{\Theta}(\tau)_j - \Theta(\tau)_j\|_2 \leq \max_{j\in H \text{ or } j+p\in H}\sup_{\tau\in\mathbb{T}}\max\{2\|\hat{B}_j(\tau)-B_j(\tau)\|_2, \|\hat{B}_j(\tau)-B_j(\tau)\|_2+\|\hat{A}_j(\tau)-A_j(\tau)\|_2\},$$

$$\max_{j\in H}\sup_{\tau\in\mathbb{T}}\|\hat{\Theta}(\tau)_j\|_1 \leq \max_{j\in H \text{ or } j+p\in H}\sup_{\tau\in\mathbb{T}}\max\{2\|\hat{B}_j(\tau)\|_1, \|\hat{B}_j(\tau)\|_1+\|\hat{A}_j(\tau)\|_1\}.$$

Combine the two cases, I have proved the first 3 inequalities in Lemma 2.

We now consider $\max_{j\in H}\sup_{\tau\in\mathbb{T}}\|\hat{\Theta}(\tau)_j'\widehat{\boldsymbol{\Sigma}}(\tau)-e_j'\|_\infty$.

$$\begin{aligned}
&\max_{j\in H\cap j\leq p}\sup_{\tau\in\mathbb{T}}\|\hat{\Theta}(\tau)_j'\widehat{\boldsymbol{\Sigma}}(\tau)-e_j'\|_\infty \\
&=\max_{j\in H\cap j\leq p}\sup_{\tau\in\mathbb{T}}\|\begin{bmatrix}\hat{B}(\tau)_j & -\hat{B}(\tau)_j\end{bmatrix}\begin{bmatrix}\widehat{\boldsymbol{M}} & \widehat{\boldsymbol{M}}(\tau)\\ \widehat{\boldsymbol{M}}(\tau) & \widehat{\boldsymbol{M}}(\tau)\end{bmatrix}-e_j'\|_\infty \\
&=\max_{j\in H\cap j\leq p}\sup_{\tau\in\mathbb{T}}\|\begin{bmatrix}\hat{B}(\tau)_j\widehat{\boldsymbol{N}}(\tau) & 0\end{bmatrix}-e_j'\|_\infty \leq \max_{j\in H\cap j\leq p}\sup_{\tau\in\mathbb{T}}\|\hat{B}(\tau)_j'\widehat{\boldsymbol{N}}(\tau)-\tilde{e}_j'\|_\infty \\
&\leq \max_{j\in H\cap j\leq p}\sup_{\tau\in\mathbb{T}}\frac{\lambda_{node}}{\widehat{\tilde{z}}_j(\tau)^2}\max_{j+p\in H} \\
&\quad\sup_{\tau\in\mathbb{T}}\|\hat{\Theta}(\tau)_j'\widehat{\boldsymbol{\Sigma}}(\tau)-e_j'\|_\infty \\
&=\max_{j+p\in H}\sup_{\tau\in\mathbb{T}}\|\begin{bmatrix}-\hat{B}(\tau)_j & \hat{B}(\tau)_j+\hat{A}(\tau)_j\end{bmatrix}\begin{bmatrix}\widehat{\boldsymbol{M}} & \widehat{\boldsymbol{M}}(\tau)\widehat{\boldsymbol{M}}(\tau) & \widehat{\boldsymbol{M}}(\tau)\end{bmatrix}-e_j'\|_\infty \\
&=\max_{j+p\in H}\sup_{\tau\in\mathbb{T}}\|\begin{bmatrix}\hat{A}(\tau)_j\widehat{\boldsymbol{M}}(\tau)-\hat{B}(\tau)_j\widehat{\boldsymbol{N}}(\tau) & \hat{A}(\tau)_j\widehat{\boldsymbol{M}}(\tau)\end{bmatrix}-\begin{bmatrix}0 & \tilde{e}_j'\end{bmatrix}\|_\infty \\
&\leq \max_{j+p\in H}\sup_{\tau\in\mathbb{T}}\max\{\|\hat{A}(\tau)_j'\widehat{\boldsymbol{M}}(\tau)-\tilde{e}_j'\|_\infty+\|\hat{B}(\tau)_j'\widehat{\boldsymbol{N}}(\tau)-\tilde{e}_j'\|_\infty, \|\hat{A}(\tau)_j'\widehat{\boldsymbol{M}}(\tau)-\tilde{e}_j'\|_\infty\} \\
&\leq \max_{j+p\in H}\sup_{\tau\in\mathbb{T}}\frac{\lambda_{node}}{\widehat{z}_j(\tau)^2}+\frac{\lambda_{node}}{\widehat{\tilde{z}}_j(\tau)^2}. \qquad \text{(A.4.11)}
\end{aligned}$$

□

A.5 Proofs for Theorem 3 for Case I. No Threshold.

This section explores the case where there is no threshold effect, i.e.the true model is linear.

To show that the ratio

$$t=\frac{\sqrt{n}g'(\hat{a}(\hat{\tau})-\alpha_0)}{\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}} \qquad \text{(A.5.1)}$$

is asymptotically standard normal. First, by rewriting (1.4.8),

$$t = t_1 + t_2,$$

where

$$t_1 = \frac{g'\widehat{\Theta}(\hat{\tau})\boldsymbol{X}'(\hat{\tau})U/n^{1/2}}{\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}} \text{ and}$$
$$t_2 = \frac{g'\Delta(\hat{\tau})}{\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}}$$

It suffices to show that t_1 is asymptotically standard normal and $t_2 = o_p(1)$.

Lemma 15. *Suppose that Assumption1,2,5 and 6 be satisfied, conditional on events* $\mathbb{A}_1$,$\mathbb{A}_2$, $\mathbb{A}_3$, $\mathbb{A}_4$, $\mathbb{A}_5$, $g'\Delta(\hat{\tau}) = O_p(\frac{s_0\sqrt{h}\log p}{\sqrt{n}})$.

Proof. By holder's inequality, Theorem 1, and Lemma 2

$$\begin{aligned}
g'\Delta(\hat{\tau}) &\leq \max_{j\in H}|\Delta_j(\hat{\tau})|\sum_{j\in H}|g_j| \\
&= \max_{j\in H}|\left(\widehat{\Theta}_j(\hat{\tau})\hat{\Sigma}(\hat{\tau}) - \tilde{e}_j'\right)\sqrt{n}(\hat{\alpha}(\hat{\tau}) - \alpha_0)|\sum_{j\in H}|g_j| \\
&\leq \max_{j\in H}|\left(\widehat{\Theta}_j(\hat{\tau})\hat{\Sigma}(\hat{\tau}) - \tilde{e}_j'\right)\sqrt{n}(\hat{\alpha}(\hat{\tau}) - \alpha_0)|\sum_{j\in H}|g_j| \\
&\leq \max_{1\leq j\leq 2p}\|\widehat{\Theta}_j(\hat{\tau})\hat{\Sigma}(\hat{\tau}) - \tilde{e}_j'\|_\infty\sqrt{n}\|\hat{\alpha}(\hat{\tau}) - \alpha_0\|_1\sum_{j\in H}|g_j| \\
&\leq C\left(\frac{\lambda_{node}}{\hat{z}_1{}^2(\hat{\tau})_j} + \frac{\lambda_{node}}{\hat{z}_2{}^2(\hat{\tau})_j}\right)\cdot\sqrt{n}\cdot\lambda s_0\sqrt{h} \\
&= O_p(\frac{s_0\sqrt{h}\log p}{\sqrt{n}})
\end{aligned}$$

□

Lemma 16. *Suppose that Assumption1 to 6 be satisfied, then*

$$\max_{1\leq k,l,j\leq p}|\frac{1}{n}\sum_{i=1}^{n}(X_i^{(k)}X_i^{(l)}X_i^{(j)})^2 - E\left[(X_i^{(k)}X_i^{(l)}X_i^{(j)})^2\right]| = O_p\left(\frac{\sqrt{\log p}}{\sqrt{n}}\right)$$

$$\max_{1\le k,l\le p}|\frac{1}{n}\sum_{i=1}^{n}(X_i^{(k)}X_i^{(l)}U_i)^2-E\left[(X_i^{(k)}X_i^{(l)}U_i)^2\right]|=O_p\left(\frac{\sqrt{\log p}}{\sqrt{n}}\right)$$

$$\max_{1\le l,k\le 2p}\sup_{\tau\in\mathbb{T}}|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i^{(k)}(\tau)\boldsymbol{X}_i^{(l)}(\tau)U_i^2-E\left[\boldsymbol{X}_i^{(k)}(\tau)\boldsymbol{X}_i^{(l)}(\tau)U_i^2\right]|=O_p\left(\frac{\sqrt{\log p}}{\sqrt{n}}\right)$$

Proof. Apply Lemma E.1 and E.2 of Chernozhukov et al. (2017) under Assumption 6 (ii) and (v), by arguments exactly parallel to those in proof of Lemma 6, and its proof therefore omitted. □

Lemma 17. *Suppose that Assumption1 to 6 be satisfied, conditional on events* $\mathbb{A}_1$ *,*$\mathbb{A}_2$*,* $\mathbb{A}_3$*,* $\mathbb{A}_4$ *and* $\mathbb{A}_5$*, then*

$$|g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g-g'\Theta(\hat{\tau})\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g|=O_p\left(h\bar{s}\sqrt{s_0^3}\sqrt{\frac{\log p}{n}}\right).$$

Proof of Lemma 17.

Recall for no-threshold case $\Sigma(\hat{\tau})_{xu}=E\left[\boldsymbol{X}_i(\hat{\tau})\boldsymbol{X}_i'(\hat{\tau})U_i^2\right]=E\left[\boldsymbol{X}_i(\hat{\tau})\boldsymbol{X}_i'(\hat{\tau})\right]E\left[U_i^2\right]$,

$$\hat{U}_i(\hat{\tau})=Y_i-\boldsymbol{X}_i'(\hat{\tau})\hat{\alpha}(\hat{\tau})=U_i+\boldsymbol{X}_i'(\hat{\tau})\alpha_0-\boldsymbol{X}_i'(\hat{\tau})\hat{\alpha}(\hat{\tau}),$$

$$\hat{\Sigma}(\hat{\tau})_{xu}=\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i(\hat{\tau})\boldsymbol{X}_i'(\hat{\tau})\hat{U}(\hat{\tau})_i^2,$$

and set $\tilde{\Sigma}(\hat{\tau})_{xu}=\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i(\hat{\tau})\boldsymbol{X}_i'(\hat{\tau})U_i^2$

We first show that

$$\begin{aligned}
&|g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g-g'\Theta(\hat{\tau})\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g|\\
=\;&|g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g-g'\widehat{\Theta}(\hat{\tau})\tilde{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g+g'\widehat{\Theta}(\hat{\tau})\tilde{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g-g'\Theta(\hat{\tau})\tilde{\Sigma}(\hat{\tau})_{xu}\Theta(\hat{\tau})'g\\
&+g'\Theta(\hat{\tau})\tilde{\Sigma}(\hat{\tau})_{xu}\Theta(\hat{\tau})'g-g'\Theta(\hat{\tau})\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g|\\
\le\;&|g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g-g'\widehat{\Theta}(\hat{\tau})\tilde{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g|+|g'\widehat{\Theta}(\hat{\tau})\tilde{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g-g'\Theta(\hat{\tau})\tilde{\Sigma}(\hat{\tau})_{xu}\Theta(\hat{\tau})'g|\\
&+|g'\Theta(\hat{\tau})\tilde{\Sigma}(\hat{\tau})_{xu}\Theta(\hat{\tau})'g-g'\Theta(\hat{\tau})\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g|
\end{aligned}$$

To prove this lemma,I need prove the followings

$$|g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g - g'\widehat{\Theta}(\hat{\tau})\tilde{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g| = o_p(1)$$

$$|g'\widehat{\Theta}(\hat{\tau})\tilde{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g - g'\widehat{\Theta}(\hat{\tau})\Sigma(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g| = o_p(1)$$

$$|g'\widehat{\Theta}(\hat{\tau})\Sigma(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g - g'\Theta(\hat{\tau})\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g| = o_p(1)$$

Step 1.

$$\begin{aligned}
& |g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g - g'\widehat{\Theta}(\hat{\tau})\tilde{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g| \\
\leq \quad & |g'\widehat{\Theta}(\hat{\tau})\left(\hat{\Sigma}(\hat{\tau})_{xu} - \tilde{\Sigma}(\hat{\tau})_{xu}\right)\widehat{\Theta}(\hat{\tau})'g| \\
\leq \quad & ||g'\widehat{\Theta}(\hat{\tau})||_1^2 ||\hat{\Sigma}(\hat{\tau})_{xu} - \tilde{\Sigma}(\hat{\tau})_{xu}||_\infty
\end{aligned}$$

$$
\begin{aligned}
&\hat{\Sigma}(\hat{\tau})_{xu} - \tilde{\Sigma}(\hat{\tau})_{xu} \\
=&\frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{X}_i(\hat{\tau})\boldsymbol{X}_i'(\hat{\tau})\hat{U}_i^2(\hat{\tau}) - \boldsymbol{X}_i(\hat{\tau})\boldsymbol{X}_i'(\hat{\tau})U_i^2\right) \\
=&\frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{X}_i(\hat{\tau})\boldsymbol{X}_i'(\hat{\tau})(U_i + \boldsymbol{X}_i'(\hat{\tau})\alpha_0 - \boldsymbol{X}_i'(\hat{\tau})\hat{\alpha}(\hat{\tau}))^2 - \boldsymbol{X}_i(\hat{\tau})\boldsymbol{X}_i'(\hat{\tau})U_i^2\right) \\
=&\frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{X}_i(\hat{\tau})\boldsymbol{X}_i'(\hat{\tau})\alpha_0'\boldsymbol{X}_i(\hat{\tau})\boldsymbol{X}_i'(\hat{\tau})\alpha_0\right) \\
+&\frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{X}_i(\hat{\tau})\boldsymbol{X}_i'(\hat{\tau})\hat{\alpha}'(\hat{\tau})\boldsymbol{X}_i(\hat{\tau})\boldsymbol{X}_i'(\hat{\tau})\hat{\alpha}(\hat{\tau})\right) \\
-&\frac{2}{n}\sum_{i=1}^{n}\left(\boldsymbol{X}_i(\hat{\tau})\boldsymbol{X}_i'(\hat{\tau})\alpha_0'\boldsymbol{X}_i(\hat{\tau})\boldsymbol{X}_i'(\hat{\tau})\hat{\alpha}(\hat{\tau})\right) \\
+&\frac{2}{n}\sum_{i=1}^{n}\left(\boldsymbol{X}_i(\hat{\tau})\boldsymbol{X}_i'(\hat{\tau})\alpha_0'\boldsymbol{X}_i(\hat{\tau})U_i\right) \\
-&\frac{2}{n}\sum_{i=1}^{n}\left(\boldsymbol{X}_i(\hat{\tau})\boldsymbol{X}_i'(\hat{\tau})\hat{\alpha}(\hat{\tau})'\boldsymbol{X}_i(\hat{\tau})U_i\right) \\
=&\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i(\hat{\tau})\boldsymbol{X}_i'(\hat{\tau})\alpha_0\boldsymbol{X}_i(\hat{\tau})\boldsymbol{X}_i'(\hat{\tau})(\alpha_0 - \hat{\alpha}(\hat{\tau})) \\
+&\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i(\hat{\tau})\boldsymbol{X}_i'(\hat{\tau})\left(\hat{\alpha}'(\hat{\tau}) - \alpha_0'\right)\boldsymbol{X}_i(\hat{\tau})\boldsymbol{X}_i'(\hat{\tau})\hat{\alpha}(\hat{\tau}) \\
+&\frac{2}{n}\sum_{i=1}^{n}\boldsymbol{X}_i(\hat{\tau})\boldsymbol{X}_i'(\hat{\tau})\left(\alpha_0' - \hat{\alpha}(\hat{\tau})'\right)\boldsymbol{X}_i(\hat{\tau})U_i
\end{aligned}
$$

Recall Lemma 16,

$$\max_{1\le k,l,j\le p} |\frac{1}{n}\sum_{i=1}^{n}(X_i^{(k)}X_i^{(l)}X_i^{(j)})^2 - E\left[(X_i^{(k)}X_i^{(l)}X_i^{(j)})^2\right]| = O_p\left(\frac{\sqrt{\log p}}{\sqrt{n}}\right)$$

$$\max_{1\le k,l\le p} |\frac{1}{n}\sum_{i=1}^{n}(X_i^{(k)}X_i^{(l)}U_i)^2 - E\left[(X_i^{(k)}X_i^{(l)}U_i)^2\right]| = O_p\left(\frac{\sqrt{\log p}}{\sqrt{n}}\right)$$

Applying Theorem 1,

$$\|\hat{\alpha}(\hat{\tau})\|_1 \le \|\alpha_0\|_1 + O_p\left(s_0\sqrt{\frac{\log p}{n}}\right)$$

$$\|\boldsymbol{X}(\hat{\tau})\hat{\alpha}(\hat{\tau})-\boldsymbol{X}(\hat{\tau})\alpha_0\|_n = O_p\left(\sqrt{s_0}\sqrt{\frac{\log p}{n}}\right)$$

$$\|\alpha_0\|_1 = O_p(s_0)$$

By Cauchy-Schwarz inequality and holder's inequality

$$\begin{aligned}
&\max_{1\le k,l\le 2p}|\tfrac{1}{n}\textstyle\sum_{i=1}^n\left(\boldsymbol{X}_i^{(k)}(\hat{\tau})\boldsymbol{X}_i^{(l)}(\hat{\tau})\alpha_0'\boldsymbol{X}_i(\hat{\tau})\left(\boldsymbol{X}_i'(\hat{\tau})\alpha_0-\boldsymbol{X}_i'(\hat{\tau})\hat{\alpha}(\hat{\tau})\right)\right)| \\
\le\ & \sqrt{\max_{1\le k,l\le 2p}\max_{1\le i\le n}\tfrac{1}{n}\textstyle\sum_{i=1}^n(\boldsymbol{X}_i^{(k)}(\hat{\tau})\boldsymbol{X}_i^{(l)}(\hat{\tau}))^2\left(\boldsymbol{X}_i'(\hat{\tau})\alpha_0\right)^2}\|\boldsymbol{X}'(\hat{\tau})\alpha_0-\boldsymbol{X}(\hat{\tau})\hat{\alpha}(\hat{\tau})\|_n \\
\le\ & \sqrt{\max_{1\le k,l\le 2p}\max_{1\le i\le n}\tfrac{1}{n}\textstyle\sum_{i=1}^n(\boldsymbol{X}_i^{(k)}(\hat{\tau})\boldsymbol{X}_i^{(l)}(\hat{\tau}))^2\left(\max_{1\le k\le 2p}\boldsymbol{X}_i^{(k)}(\tau_0)\right)^2\|\alpha_0\|_1^2}\|\boldsymbol{X}(\tau_0)\alpha_0-\boldsymbol{X}(\hat{\tau})\hat{\alpha}(\hat{\tau})\|_n \\
\le\ & \sqrt{\max_{1\le k,l,j\le p}\tfrac{1}{n}\textstyle\sum_{i=1}^n(X_i^{(k)}X_i^{(l)}X_i^{(j)})^2\|\alpha_0\|_1^2\cdot\mathbf{1}(Q_i<\hat{\tau})}\|\boldsymbol{X}(\hat{\tau})\hat{\alpha}(\hat{\tau})-\boldsymbol{X}(\hat{\tau})\alpha_0\|_n \\
\le\ & O_p\left(\sqrt{s_0^3}\sqrt{\tfrac{\log p}{n}}\right),
\end{aligned}$$

$$\begin{aligned}
&\max_{1\le k,l\le 2p}|\tfrac{1}{n}\textstyle\sum_{i=1}^n\left(\boldsymbol{X}_i(\hat{\tau})\boldsymbol{X}_i'(\hat{\tau})\left(\hat{\alpha}'(\hat{\tau})-\alpha_0'\right)\boldsymbol{X}_i(\hat{\tau})\boldsymbol{X}_i'(\hat{\tau})\hat{\alpha}(\hat{\tau})\right) \\
\le\ & \max_{1\le k,l\le 2p}\max_{1\le i\le n}\sqrt{\tfrac{1}{n}\textstyle\sum_{i=1}^n(\boldsymbol{X}_i^{(k)}(\hat{\tau})\boldsymbol{X}_i^{(l)}(\hat{\tau}))^2\left(\hat{\alpha}'(\hat{\tau})\boldsymbol{X}_i(\hat{\tau})\right)^2}\|\boldsymbol{X}(\hat{\tau})\hat{\alpha}(\hat{\tau})-\boldsymbol{X}(\hat{\tau})\alpha_0\|_n \\
\le\ & \sqrt{\max_{1\le k,l,j\le p}\tfrac{1}{n}\textstyle\sum_{i=1}^n(X_i^{(k)}X_i^{(l)}X_i^{(j)})^2\|\hat{\alpha}(\hat{\tau})\|_1^2|1(Q_i<\hat{\tau})}\|\boldsymbol{X}(\hat{\tau})\hat{\alpha}(\hat{\tau})-\boldsymbol{X}(\hat{\tau})\alpha_0\|_n \\
\le\ & O_p\left(\sqrt{s_0^3}\sqrt{\tfrac{\log p}{n}}\right)
\end{aligned}$$

$$\begin{aligned}
&\max_{1\le k,l\le 2p}|\tfrac{2}{n}\textstyle\sum_{i=1}^n\left(\boldsymbol{X}_i'(\hat{\tau})\alpha_0-\boldsymbol{X}_i'(\hat{\tau})\hat{\alpha}(\hat{\tau})\right)(\boldsymbol{X}_i^{(k)}(\hat{\tau})\boldsymbol{X}_i^{(l)}(\hat{\tau}))U_i| \\
\le\ & 2\sqrt{\max_{1\le k,l\le p}\tfrac{1}{n}\textstyle\sum_{i=1}^n(X_i^{(k)}X_i^{(l)}U_i)^2\cdot 1(Q_i<\hat{\tau})}\|\boldsymbol{X}(\hat{\tau})\hat{\alpha}(\hat{\tau})-\boldsymbol{X}(\hat{\tau})\alpha_0\|_n \\
\le\ & O_p\left(\sqrt{s_0}\sqrt{\tfrac{\log p}{n}}\right)
\end{aligned}$$

Hence,

$$\|\hat{\Sigma}(\hat{\tau})_{xu}-\tilde{\Sigma}(\hat{\tau})_{xu}\|_\infty = O_p\left(\sqrt{s_0^3}\sqrt{\frac{\log p}{n}}\right)$$

$$
\begin{aligned}
& |g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g - g'\widehat{\Theta}(\hat{\tau})\tilde{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g| \\
\leq & \quad \|g'\widehat{\Theta}(\hat{\tau})\|_1^2 \|\hat{\Sigma}(\hat{\tau})_{xu} - \tilde{\Sigma}(\hat{\tau})_{xu}\|_\infty \\
\leq & \left(\sum_{j\in H}|g_j| \max_{j\in H}\sup_{\tau\in\mathbb{T}}\|\widehat{\Theta}(\hat{\tau})\|_1\right)^2 \|\hat{\Sigma}(\hat{\tau})_{xu} - \tilde{\Sigma}(\hat{\tau})_{xu}\|_\infty \\
= & \quad O_p(h\bar{s})\, O_p\left(\sqrt{s_0^3}\sqrt{\tfrac{\log p}{n}}\right) = O_p\left(h\bar{s}\sqrt{s_0^3}\sqrt{\tfrac{\log p}{n}}\right)
\end{aligned}
$$

Step 2.Next, I show that

$$
|g'\widehat{\Theta}(\hat{\tau})\tilde{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g - g'\widehat{\Theta}(\hat{\tau})\Sigma(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g| = o_p(1)
$$

Note that

$$
\begin{aligned}
& \tilde{\Sigma}(\hat{\tau})_{xu} - \Sigma(\hat{\tau})_{xu} \\
= & \quad \tfrac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i(\hat{\tau})\boldsymbol{X}_i'(\hat{\tau})U_i^2 - E\left[\boldsymbol{X}_i(\hat{\tau})\boldsymbol{X}_i'(\hat{\tau})U_i^2\right]
\end{aligned}
$$

Recall Lemma 16,

$$
\max_{1\leq l,k\leq 2p} |\frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i^{(k)}(\hat{\tau})\boldsymbol{X}_i^{(l)}(\hat{\tau})U_i^2 - E\left[\boldsymbol{X}_i^{(k)}(\hat{\tau})\boldsymbol{X}_i^{(l)}(\hat{\tau})U_i^2\right]| = O_p\left(\frac{\sqrt{\log p}}{\sqrt{n}}\right)
$$

Therefore

$$
\begin{aligned}
& |g'\widehat{\Theta}(\hat{\tau})\tilde{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g - g'\widehat{\Theta}(\hat{\tau})\Sigma(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g| \\
\leq & \quad |g'\widehat{\Theta}(\hat{\tau})\left(\tilde{\Sigma}(\hat{\tau})_{xu} - \Sigma(\hat{\tau})_{xu}\right)\widehat{\Theta}(\hat{\tau})'g| \\
\leq & \quad \|g'\widehat{\Theta}(\hat{\tau})\|_1^2 \|\tilde{\Sigma}(\hat{\tau})_{xu} - \Sigma(\hat{\tau})_{xu}\|_\infty \\
\leq & \left(\sum_{j\in H}|g_j| \max_{j\in H}\sup_{\tau\in\mathbb{T}}\|\widehat{\Theta}_j(\hat{\tau})\|_1\right)^2 \|\tilde{\Sigma}(\hat{\tau})_{xu} - \Sigma(\hat{\tau})_{xu}\|_\infty \\
\leq & \quad O_p(h\bar{s})\, O_p\left(\sqrt{\tfrac{\log p}{n}}\right) = O_p\left(h\bar{s}\sqrt{\tfrac{\log p}{n}}\right)
\end{aligned}
$$

Step 3.Next, I show that

$$|g'\widehat{\Theta}(\hat{\tau})\Sigma(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g - g'\Theta(\hat{\tau})\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g| = o_p(1)$$

By Lemma 6.1 in van de Geer et al. (2014)

$$\begin{aligned}
& |g'\widehat{\Theta}(\hat{\tau})\Sigma(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g - g'\Theta(\hat{\tau})\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g| \\
\leq \; & \|\Sigma(\hat{\tau})_{xu}\|_\infty \|\left(\widehat{\Theta}(\hat{\tau})-\Theta(\hat{\tau})\right)' g\|_1^2 + 2\|\left(\widehat{\Theta}(\hat{\tau})-\Theta(\hat{\tau})\right)' g\|_2 \|\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g\|_2 \\
= \; & \|\Sigma(\hat{\tau})_{xu}\|_\infty \|\left(\widehat{\Theta}(\hat{\tau})-\Theta(\hat{\tau})\right)' g\|_1^2 + 2\tilde{\kappa}(\bar{s}, c_0, \mathbb{T}, \Sigma_{xu})\|\left(\widehat{\Theta}(\hat{\tau})-\Theta(\hat{\tau})\right)' g\|_2 \|\Theta(\hat{\tau})'g\|_2 \\
\leq \; & \|\Sigma(\hat{\tau})_{xu}\|_\infty \|\left(\widehat{\Theta}(\hat{\tau})-\Theta(\hat{\tau})\right)' g\|_1^2 + 2\tilde{\kappa}(\bar{s}, c_0, \mathbb{T}, \Sigma_{xu})\|\left(\widehat{\Theta}(\hat{\tau})-\Theta(\hat{\tau})\right)' g\|_2 \tilde{\kappa}(\bar{s}, c_0, \mathbb{T}, \Theta)\|g\|_2
\end{aligned}$$

As $\|\Sigma(\hat{\tau})_{xu}\|_\infty = \max_{1\leq l,k\leq 2p} E\left[\boldsymbol{X}_i^{(k)}(\hat{\tau})\boldsymbol{X}_i^{(l)}(\hat{\tau})u_i^2\right]$, $\tilde{\kappa}(\bar{s}, c_0, \mathbb{T}, \Sigma_{xu})$ and $\tilde{\kappa}(\bar{s}, c_0, \mathbb{T}, \Theta)$ are assumed bounded from Assumption 6,

$$\begin{aligned}
& \|\left(\widehat{\Theta}(\hat{\tau})-\Theta(\hat{\tau})\right)' g\|_1 \\
= & \sum_{j\in H}\left(|g_j|\|\Theta_j(\hat{\tau})-\Theta_j(\tau_0)\|_1\right) \\
\leq & \sum_{j\in H}|g_j| \sup_{\tau\in\widehat{\mathbb{T}}} \max_{j\in H} \|\Theta_j(\tau)-\Theta_j(\tau_0)\|_1 \\
\leq & \sqrt{h} \sup_{\tau\in\widehat{\mathbb{T}}} \max_{j\in H} \|\Theta_j(\tau)-\Theta_j(\tau_0)\|_1 \\
= & O_p\left(\sqrt{h}\bar{s}\sqrt{\frac{\log p}{n}}\right)
\end{aligned}$$

$$\begin{aligned}
& \|\left(\widehat{\Theta}(\hat{\tau})-\Theta(\hat{\tau})\right)' g\|_2 \\
= & \|\sum_{j\in H}\left(\Theta_j(\hat{\tau})-\Theta_j(\tau_0)\right)|g_j|\|_2 \\
\leq & \max_{j\in H}\|\Theta_j(\hat{\tau})-\Theta_j(\tau_0)\|_2 \sum_{j\in H}|g_j| \\
\leq & \sqrt{h} \sup_{\tau\in\widehat{\mathbb{T}}} \max_{j\in H} \|\Theta_j(\tau)-\Theta_j(\tau_0)\|_2 \\
= & O_p\left(\sqrt{h\bar{s}}\sqrt{\frac{\log p}{n}}\right)
\end{aligned}$$

Furthermore,

$$
\begin{aligned}
& |g'\widehat{\Theta}(\hat{\tau})\Sigma(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g - g'\Theta(\hat{\tau})\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g| \\
\leq & \ \|\Sigma(\hat{\tau})_{xu}\|_\infty \|\left(\widehat{\Theta}(\hat{\tau})-\Theta(\hat{\tau})\right)' g\|_1^2 + 2\tilde{\kappa}(\bar{s}, c_0, \mathbb{T}, \Sigma_{xu})\|\left(\widehat{\Theta}(\hat{\tau})-\Theta(\hat{\tau})\right)' g\|_2 \tilde{\kappa}(\bar{s}, c_0, \mathbb{T}, \Theta)\|g\|_2 \\
\leq & \ O_p\left(\sqrt{h}\bar{s}\sqrt{\frac{\log p}{n}}\right)^2 + O_p\left(\sqrt{h}\bar{s}\sqrt{\frac{\log p}{n}}\right) \\
= & \ O_p\left(\sqrt{h}\bar{s}\sqrt{\frac{\log p}{n}}\right)
\end{aligned}
$$

Finally, by Assumption 6 (ii),

$$
\begin{aligned}
& |\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})' - \Theta(\hat{\tau})\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'| \\
= & \ O_p\left(h\bar{s}\sqrt{s_0^3}\sqrt{\frac{\log p}{n}}\right) + O_p\left(h\bar{s}\sqrt{\frac{\log p}{n}}\right) + O_p\left(\sqrt{h}\bar{s}\sqrt{\frac{\log p}{n}}\right) = O_p\left(h\bar{s}\sqrt{s_0^3}\sqrt{\frac{\log p}{n}}\right)
\end{aligned}
$$

□

Proof of Theorem 3 Case I: no threshold. **Step 1**.

Step 1.1) Given that τ_0 is undefined and unknown in the current setup, it is necessary to show the asymptotic standard normality of $t_1'(\tau) = \frac{g'\Theta(\tau)\boldsymbol{X}'(\tau)U/n^{1/2}}{\sqrt{g'\Theta(\tau)\Sigma(\tau)_{xu}\Theta(\tau)'g}}$ uniformly over $\tau \in \mathbb{T}$. Subsequently, for any $\widehat{\tau}$ obtained from (1.2.4), I need to show get $t_1'(\widehat{\tau})$ and t_1 are asymptotically equivalent.

Note that, using $E(U_i|X_i) = 0$ for all $i = 1, ..., n$, I obtain

$$
E\left[t_1'(\tau)\right] = E\left[\frac{g'\Theta(\tau)\sum_{i=1}^n \boldsymbol{X}_i(\tau)U_i/n^{1/2}}{\sqrt{g'\Theta(\tau)\Sigma(\tau)_{xu}\Theta(\tau)'g}}\right] = 0, \tag{A.5.2}
$$

and

$$
E\left[t_1'(\tau)\right]^2 = E\left[\frac{g'\Theta(\tau_0)\sum_{i=1}^n \boldsymbol{X}_i(\tau)U_i/n^{1/2}}{\sqrt{g'\Theta(\tau)\Sigma(\tau)_{xu}\Theta(\tau)'g}}\right]^2 = 1.
$$

Hence, to use Lyapounov's central limit theorem, I check the conditions for a sequence of independent random variables, it suffices to show that for some $\varepsilon > 0$

$$\frac{\sum_{i=1}^n E|g'\Theta(\tau)\boldsymbol{X}_i'(\tau)U_i/n^{1/2}|^{2+\varepsilon}}{(g'\Theta(\tau)\Sigma(\tau)_{xu}\Theta(\tau)'g)^{1+\varepsilon/2}} \to 0.$$

Let $\tilde{S}(\tau) = \cup_{j\in H} S_j(\tau)$, then the cardinality $\sup_{\tau\in\mathbb{T}} |\tilde{S}(\tau)| = p \wedge h\bar{s}$.

$$\begin{aligned}
E\left|g'\Theta(\tau)\boldsymbol{X}_i'(\tau)U_i/n^{1/2}\right|^{2+\varepsilon} &\le E\left|\left[\|g'\Theta(\tau)/n^{1/2}\|_1 \max_{j\in\tilde{S}(\tau)}\left(\boldsymbol{X}_i^{(j)}(\tau)U_i\right)\right]\right|^{2+\varepsilon} \\
&\le E\left[\|g'\Theta(\tau)/n^{1/2}\|_1^{2+\varepsilon} \max_{j\in\tilde{S}(\tau)}\left|\boldsymbol{X}_i^{(j)}(\tau)U_i\right|^{2+\varepsilon}\right] \\
&\le \|g'\Theta(\tau)/n^{1/2}\|_1^{2+\varepsilon} E\left[\max_{j\in\tilde{S}(\tau)}\left|\boldsymbol{X}_i^{(j)}(\tau)U_i\right|^{2+\varepsilon}\right] \\
&\le \|g'\Theta(\tau)/n^{1/2}\|_1^{2+\varepsilon} E\left[\sum_{j\in\tilde{S}(\tau)}\left|\boldsymbol{X}_i^{(j)}(\tau)U_i\right|^{2+\varepsilon}\right] \\
&\le \|g'\Theta(\tau)/n^{1/2}\|_1^{2+\varepsilon}(p\wedge h\bar{s}) \max_{j\in\tilde{S}(\tau)} E\left[\left|\boldsymbol{X}_i^{(j)}(\tau)U_i\right|^{2+\varepsilon}\right] \\
&\le \|g'\Theta(\tau)/n^{1/2}\|_1^{2+\varepsilon}(p\wedge h\bar{s}) \max_{1\le j\le p} E\left[\left|X_i^{(j)}U_i\right|^{2+\varepsilon}\right] \\
&= O_p\left(\frac{(h\bar{s})^{2+\varepsilon/2}}{n^{1+\varepsilon/2}}\right)\max_{1\le j\le p} E\left[\left(X_i^{(j)}U_i\right)^{2+\varepsilon}\right] \wedge O_p\left(\frac{(h\bar{s})^{1+\varepsilon/2}p}{n^{1+\varepsilon/2}}\right)\max_{1\le j\le p} E\left[\left(X_i^{(j)}U_i\right)^{2+\varepsilon}\right]
\end{aligned}$$

where the 1st inequality follows from the Holder's inequality.

By Cauchy–Schwarz inequality

$$E\left[\left(X_i^{(j)}U_i\right)^4\right] \le E\left[\left(X_i^{(j)}\right)^4\right] E\left[(U_i)^4\right]$$

is bounded by assumption 1.

Thus take $\varepsilon = 2$, $\sum_{i=1}^n E\left|g'\Theta(\tau)\boldsymbol{X}_i(\tau)U_i/n^{1/2}\right|^4 = O_p\left(\frac{(h\bar{s})^3}{n^2}\right) \wedge O_p\left(\frac{(h\bar{s})^2 p}{n^2}\right) = o_p(1)$ by Assumption 6 (iv)

Next, I show that $g'\Theta(\tau_0)\Sigma(\tau)_{xu}\Theta(\tau)'g$ is asymptotically bounded away from zero. Clearly,

$$\begin{aligned}
g'\Theta(\tau)\Sigma(\tau)_{xu}\Theta(\tau)'g &\ge \kappa(\bar{s}, c_0, \mathbb{T}, \Sigma_{xu})\|g'\Theta(\tau)\|_2^2 \\
&\ge \kappa(\bar{s}, c_0, \mathbb{T}, \Sigma_{xu})\|g'\|_2^2 \kappa(\bar{s}, c_0, \mathbb{T}, \Theta)^2 \qquad \text{(A.5.3)} \\
&= \kappa(\bar{s}, c_0, \mathbb{T}, \Sigma_{xu})\kappa(\bar{s}, c_0, \mathbb{T}, \Theta)^2,
\end{aligned}$$

which is bounded away from zero since $\kappa(\bar{s}, c_0, \mathbb{T}, \Sigma_{xu})$ and $\kappa(\bar{s}, c_0, \mathbb{T}, \Theta)$ are bounded away from zero by Assumption 6 (iv). Hence, the Lyapunov condition is satisfied and $\forall \tau \in \mathbb{T}$, $t_1'(\tau)$

converges in distribution to a standard normal.

Step 1.2).

Let

$$t_1'' = \frac{g'\Theta(\widehat{\tau})\boldsymbol{X}'(\widehat{\tau})U/n^{1/2}}{\sqrt{g'\widehat{\Theta}(\widehat{\tau})\widehat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})g}}$$

$$|g'\widehat{\Theta}(\widehat{\tau})\boldsymbol{X}'(\widehat{\tau})U/n^{1/2} - g'\Theta(\widehat{\tau})\boldsymbol{X}'(\widehat{\tau})U/n^{1/2}|$$

$$\leq \|g'\left(\widehat{\Theta}(\widehat{\tau}) - \Theta(\widehat{\tau})\right)\|_1 \|\boldsymbol{X}(\widehat{\tau})U/n^{1/2}\|_\infty$$

Conditional on $\mathbb{A}_1$, $\mathbb{A}_2$, $\mathbb{A}_3$ and $\mathbb{A}_4$ and by Lemma 2

$$= O_p(\sqrt{h}\bar{s}\frac{\sqrt{\log p}}{\sqrt{n}})O_p(\sqrt{\log p}) = O_p(\sqrt{h}\bar{s}\frac{\log p}{\sqrt{n}}) = o_p(1)$$

$$|t_1'' - t_1| = \frac{g'\left(\Theta(\widehat{\tau})\boldsymbol{X}'(\widehat{\tau})U/n^{1/2} - \widehat{\Theta}(\widehat{\tau})\boldsymbol{X}'(\widehat{\tau})U/n^{1/2}\right)}{\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}} = o_p(1)$$

$$\begin{aligned}
|t_1 - t_1'| =& \frac{\left(\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g} - \sqrt{g'\Theta(\hat{\tau})\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g}\right)g'\Theta(\widehat{\tau})X'(\widehat{\tau})U/n^{1/2}}{\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}\sqrt{g'\Theta(\hat{\tau})\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g}} \\
=& \frac{\left(g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g - g'\Theta(\hat{\tau})\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g\right)g'\Theta(\widehat{\tau})X'(\widehat{\tau})U/n^{1/2}}{\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}\sqrt{g'\Theta(\hat{\tau})\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g}\left(\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g} + \sqrt{g'\Theta(\hat{\tau})\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g}\right)} \\
\leq& \frac{|g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g - g'\Theta(\hat{\tau})\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g|g'\Theta(\widehat{\tau})X'(\widehat{\tau})U/n^{1/2}}{\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}\sqrt{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}\left(\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g} + \sqrt{g'\Theta(\hat{\tau})\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g}\right)} \\
\leq& \frac{O_p\left(h\sqrt{s_0^3\bar{s}^2}\sqrt{\frac{\log p}{n}}\right)O_p\left(\sqrt{h\bar{s}\log p}\right)}{\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}\sqrt{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}\left(\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g} + \sqrt{g'\Theta(\hat{\tau})\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g}\right)} \\
=& o_p(1)
\end{aligned}$$

by Lemma 17.

Then combine the above two,

$$|t_1 - t_1'| \leq o_p(1)$$

Step 2. By Lemma 15,

$$t_2 = \frac{g'\Delta(\hat{\tau})}{\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}} = o_p(1).$$

Finally, by Slutsky's theorem

$$t = o_p(1) + t_1' \xrightarrow{d} N(0,1).$$

□

A.6 Proofs for Theorem 3 for Case II. Fixed Threshold.

This section explores the case where the threshold effect is well-identified and discontinuous. To show that the ratio

$$t = \frac{\sqrt{n}g'(\hat{a}(\hat{\tau}) - \alpha_0)}{\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}} \tag{A.6.1}$$

is asymptotically standard normal. First, by rewriting (1.4.11),

$$t = t_1 + t_2,$$

where $t_1 = \frac{g'\widehat{\Theta}(\tau_0)\boldsymbol{X}'(\tau_0)U/n^{1/2}}{\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}}$, and

$$t_2 = \frac{g'(\hat{\Theta}(\hat{\tau})\boldsymbol{X}'(\hat{\tau})U - g'\hat{\Theta}(\tau_0)\boldsymbol{X}'(\tau_0)U)//n^{1/2} + g'\hat{\Theta}(\hat{\tau})(\boldsymbol{X}'(\hat{\tau})\boldsymbol{X}(\tau_0) - \boldsymbol{X}'(\hat{\tau})\boldsymbol{X}(\hat{\tau}))\alpha_0/n^{1/2} - g'\Delta(\hat{\tau})}{\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}}.$$

It suffices to show that t_1 is asymptotically standard normal and $t_2 = o_p(1)$.

Lemma 18. *Suppose that Assumption1 to 6 be satisfied, conditional on events* $\mathbb{A}_1$,$\mathbb{A}_2$, $\mathbb{A}_3$, $\mathbb{A}_4$, $\mathbb{A}_5$, *then* $g'\Delta(\hat{\tau}) = O_p(\frac{s_0\sqrt{h}\log p}{\sqrt{n}})$.

Proof. Recall that $\Delta(\tau) = \sqrt{n}(\hat{\Theta}(\tau)\hat{\Sigma}(\tau) - I_{2p})(\hat{\alpha}(\tau) - \alpha_0)$

Thus by holder's inequality, Lemma 2 and Theorem 2,

$$\begin{aligned}
g'\Delta(\hat{\tau}) &\leq \max_{j\in H}|\Delta_j(\hat{\tau})|\sum_{j\in H}|g_j| \\
&= \max_{j\in H}|\left(\widehat{\Theta}_j(\hat{\tau})\hat{\Sigma}(\hat{\tau})-\tilde{e}_j'\right)\sqrt{n}(\hat{\alpha}(\hat{\tau})-\alpha_0)|\sum_{j\in H}|g_j| \\
&\leq \max_{j\in H}|\left(\widehat{\Theta}_j(\hat{\tau})\hat{\Sigma}(\hat{\tau})-\tilde{e}_j'\right)\sqrt{n}(\hat{\alpha}(\hat{\tau})-\alpha_0)|\sum_{j\in H}|g_j| \\
&\leq \max_{1\leq j\leq 2p}\|\widehat{\Theta}_j(\hat{\tau})\hat{\Sigma}(\hat{\tau})-\tilde{e}_j'\|_\infty\sqrt{n}\|\hat{\alpha}(\hat{\tau})-\alpha_0\|_1\sum_{j\in H}|g_j| \\
&\leq C\left(\frac{\lambda_{node}}{\hat{z}_1{}^2(\hat{\tau})_j}+\frac{\lambda_{node}}{\hat{z}_2{}^2(\hat{\tau})_j}\right)\cdot\sqrt{n}\cdot\lambda s_0\sqrt{h} \\
&= O_p(\frac{s_0\sqrt{h}\log p}{\sqrt{n}})
\end{aligned}$$

□

The results of Lemma 18 are similar to those in Lemma 15 but the assumptions differ.

Lemma 19. *Let Assumptions 1, 2, 3, 4, 5 and 6 be satisfied and let* g *be* $2p\times 1$ *vector satisfying* $\|g\|_2=1$*. Then,conditional on events* $\mathbb{A}_1$,$\mathbb{A}_2$, $\mathbb{A}_3$, $\mathbb{A}_4$, $\mathbb{A}_5$,

$$\|g'\left(\widehat{\Theta}(\hat{\tau})-\widehat{\Theta}(\tau_0)\right)\|_1=O_p\left(\sqrt{h}\bar{s}\sqrt{\frac{\log p}{n}}\right)$$

Proof of Lemma. As Q_i are continuously distributed and $E\left[|X_i^{(j)}X_i^{(l)}||Q_i=\tau\right]$ is continuous and bounded in a neighborhood of τ_0, conditions for Lemma A.1 in Hansen (2000) hold. Then

$$\begin{aligned}
&\|\Sigma(\tau_0)-\Sigma(\widehat{\tau})\|_\infty \\
=&\|\begin{bmatrix} 0 & \boldsymbol{M}(\tau_0)-\boldsymbol{M}(\widehat{\tau}) \\ \boldsymbol{M}(\tau_0)-\boldsymbol{M}(\widehat{\tau}) & \boldsymbol{M}(\tau_0)-\boldsymbol{M}(\widehat{\tau})\end{bmatrix}\|_\infty \\
\leq&\|\boldsymbol{M}(\tau_0)-\boldsymbol{M}(\widehat{\tau})\|_\infty \\
=&\max_{1\leq j,l\leq p}E\left[|X_i^{(j)}X_i^{(l)}||1(Q_i<\tau_0)-1(Q_i<\widehat{\tau})|\right] \\
\leq&C|\tau_0-\widehat{\tau}| \\
=&O_p\left(\frac{\log p\, s_0}{n}\right)
\end{aligned}$$

where the last inequality is by Lemma A.1 in Hansen (2000) and the last line is due to Theorem 2.

Consider

$$
\begin{aligned}
& \|\Theta_j(\widehat{\tau})-\Theta_j(\tau_0)\|_1 \\
= & \quad \|\Theta_j(\widehat{\tau})\big(\Sigma_j(\tau_0)-\Sigma_j(\tau)\big)'\Theta_j(\tau_0)\|_1 \\
\leq & \quad \|\Theta_j(\widehat{\tau})\|_1\|\big(\Sigma_j(\tau_0)-\Sigma_j(\widehat{\tau})\big)'\Theta_j(\tau_0)\|_\infty \\
\leq & \quad \|\Theta_j(\widehat{\tau})\|_1\|\Theta_j(\tau_0)\|_1\|\big(\Sigma_j(\tau_0)-\Sigma_j(\widehat{\tau})\big)'\|_\infty
\end{aligned}
$$

Then using Lemma 2

$$
\begin{aligned}
& \|g'(\Theta(\widehat{\tau})-\Theta(\tau_0))\|_1 \\
= & \sum_{j\in H}\big(|g_j|\|\Theta_j(\widehat{\tau})-\Theta_j(\tau_0)\|_1\big) \\
\leq & \sum_{j\in H}|g_j|\sup_{\tau\in\mathbb{T}}\max_{j\in H}\|\Theta_j(\tau)-\Theta_j(\tau_0)\|_1 \\
\leq & \sqrt{h}\sup_{\tau\in\mathbb{T}}\max_{j\in H}\|\Theta_j(\tau)\|_1\max_{j\in H}\|\Theta_j(\tau_0)\|_1\|\big(\Sigma_j(\tau_0)-\Sigma_j(\tau)\big)'\|_\infty \\
= & O_p\left(\sqrt{h}\bar{s}s_0\frac{\log p}{n}\right)
\end{aligned}
\tag{A.6.2}
$$

Finally,

$$
\begin{aligned}
& \|g'\big(\widehat{\Theta}(\hat{\tau})-\widehat{\Theta}(\tau_0)\big)\|_1 \\
\leq & \sum_{j\in H}|g_j|\sup_{\tau\in\mathbb{T}}\max_{j\in H}\|\widehat{\Theta}_j(\tau)-\Theta_j(\tau)\|_1+\sum_{j\in H}|g_j|\sup_{\tau\in\mathbb{T}}\max_{j\in H}\|\Theta_j(\tau)-\Theta_j(\tau_0)\|_1+\sum_{j\in H}|g_j|\sup_{\tau\in\mathbb{T}}\max_{j\in H}\|\widehat{\Theta}_j(\tau_0)-\Theta_j(\tau_0)\|_1 \\
= & O_p\left(\sqrt{h}\bar{s}\sqrt{\frac{\log p}{n}}\right)+O_p\left(\sqrt{h}\bar{s}s_0\frac{\log p}{n}\right)=O_p\left(\sqrt{h}\bar{s}\sqrt{\frac{\log p}{n}}\right)
\end{aligned}
$$

since $s_0\sqrt{\frac{\log p}{n}}=o_p(1)$ by Assumption 1. □

Lemma 20. *Suppose that Assumption1 to 6 be satisfied, conditional on events* $\mathbb{A}_1$ *,*$\mathbb{A}_2$*,* $\mathbb{A}_3$*,* $\mathbb{A}_4$*,* $\mathbb{A}_5$*,*

$$
|g'\big(\hat{\Theta}(\hat{\tau})\boldsymbol{X}'(\hat{\tau})U-\hat{\Theta}(\tau_0)\boldsymbol{X}'(\tau_0)U\big)/n^{1/2}|=O_p\left(\sqrt{h}\bar{s}\frac{\log p}{\sqrt{n}}\right)+O_p\left(\frac{\sqrt{h\bar{s}s_0}\log p}{\sqrt{n}}\right)
$$

Proof of Lemma . To prove this lemma,I need prove the followings

$$|g'\left(\hat{\Theta}(\hat{\tau})-\hat{\Theta}(\tau_0)\right)\boldsymbol{X}'(\tau_0)U|/\sqrt{n}=o_p(1),$$

$$|g'\left(\hat{\Theta}(\hat{\tau})\boldsymbol{X}'(\hat{\tau})-\hat{\Theta}(\hat{\tau})\boldsymbol{X}'(\tau_0)\right)U|/\sqrt{n}=o_p(1).$$

On the event $\mathbb{A}_1$,$\mathbb{A}_3$ and $\mathbb{A}_4$

$$\|\frac{\boldsymbol{X}'(\tau_0)U}{\sqrt{n}}\|_\infty \quad \leq \quad \frac{1}{2}\sqrt{n}\mu\lambda\sqrt{C_2^2+\mu\lambda}$$

Then, by Hölder's inequality and Lemma 19

$$\begin{aligned}
&|g'\left(\hat{\Theta}(\hat{\tau})\boldsymbol{X}'(\tau_0)-\hat{\Theta}(\tau_0)\boldsymbol{X}'(\tau_0)\right)U|/\sqrt{n}\\
&\leq\|g'\left(\hat{\Theta}(\hat{\tau})-\hat{\Theta}(\tau_0)\right)\|_1\|\frac{\boldsymbol{X}'(\tau_0)U}{\sqrt{n}}\|_\infty\\
&\leq O_p\left(\sqrt{h}\bar{s}\sqrt{\frac{\log p}{n}}\right)O_p(\sqrt{\log p})\\
&=O_p\left(\sqrt{h}\bar{s}\frac{\log p}{\sqrt{n}}\right),
\end{aligned} \tag{A.6.3}$$

Considering $\|\frac{\boldsymbol{X}'(\hat{\tau})U}{\sqrt{n}}-\frac{\boldsymbol{X}'(\tau_0)U}{\sqrt{n}}\|_\infty$, by Assumption 4 (1.3.6)

$$\|\frac{\boldsymbol{X}'(\hat{\tau})U}{n}-\frac{\boldsymbol{X}'(\tau_0)U}{n}\|_\infty=O_p(\frac{\sqrt{s_0}\log p}{n}).$$

$$\begin{aligned}
&|g'\hat{\Theta}(\hat{\tau})\left(\boldsymbol{X}'(\hat{\tau})-\boldsymbol{X}'(\tau_0)\right)U|/\sqrt{n}\\
&\leq\sqrt{n}\|\frac{\boldsymbol{X}'(\hat{\tau})U}{n}-\frac{\boldsymbol{X}'(\tau_0)U}{n}\|_\infty\|g'\hat{\Theta}(\hat{\tau})\|_1\\
&\leq\sqrt{n}\|\frac{\boldsymbol{X}'(\hat{\tau})U}{n}-\frac{\boldsymbol{X}'(\tau_0)U}{n}\|_\infty\|g\|_1\max_{j\in H}\|\widehat{\Theta}_j(\hat{\tau})\|_1\\
&\leq\sqrt{n}O_p\left(\sqrt{h\bar{s}}\right)O_p\left(\frac{\sqrt{s_0}\log p}{n}\right)=O_p\left(\frac{\sqrt{h\bar{s}s_0}\log p}{\sqrt{n}}\right)
\end{aligned} \tag{A.6.4}$$

Hence, combine (A.6.3) and (A.6.4)

$$|g'\left(\hat{\Theta}(\hat{\tau})\boldsymbol{X}'(\hat{\tau})U-\hat{\Theta}(\tau_0)\boldsymbol{X}'(\tau_0)U\right)/n^{1/2}|=O_p\left(\sqrt{h}\bar{s}\frac{\log p}{\sqrt{n}}\right)+O_p\left(\frac{\sqrt{h\bar{s}s_0}\log p}{\sqrt{n}}\right)$$

□

Lemma 21. *Suppose that Assumption1 to 6 be satisfied, conditional on events* $\mathbb{A}_1$,$\mathbb{A}_2$, $\mathbb{A}_3$, $\mathbb{A}_4$, $\mathbb{A}_5$, *then*

$$|g'\hat{\Theta}(\hat{\tau})(\boldsymbol{X}'(\hat{\tau})\boldsymbol{X}(\tau_0)-\boldsymbol{X}'(\hat{\tau})\boldsymbol{X}(\hat{\tau}))\alpha_0/n^{1/2}| = O_p\left(\frac{s_0^2\sqrt{h\bar{s}}\log p}{\sqrt{n}}\right)$$

Proof of Lemma 21. There are only two cases for $\boldsymbol{X}'(\hat{\tau})\boldsymbol{X}(\tau_0)$:

$\boldsymbol{X}'(\hat{\tau})\boldsymbol{X}(\tau_0) = \boldsymbol{X}'(\tau_0)\boldsymbol{X}(\tau_0)$ or $\boldsymbol{X}'(\hat{\tau})\boldsymbol{X}(\tau_0) = \boldsymbol{X}'(\hat{\tau})\boldsymbol{X}(\hat{\tau})$, then

$$\begin{aligned}
&|g'\hat{\Theta}(\hat{\tau})(\boldsymbol{X}'(\hat{\tau})\boldsymbol{X}(\tau_0)-\boldsymbol{X}'(\hat{\tau})\boldsymbol{X}(\hat{\tau}))\alpha_0/n^{1/2}| \\
\leq &\sqrt{n}\sum_{j\in H}|g_j|\|\widehat{\Theta}_j(\hat{\tau})\|_1\|\begin{bmatrix} 0 & \widehat{\boldsymbol{M}}(\tau_0)-\widehat{\boldsymbol{M}}(\hat{\tau}) \\ 0 & \widehat{\boldsymbol{M}}(\min\{\tau_0,\hat{\tau}\})-\widehat{\boldsymbol{M}}(\hat{\tau}) \end{bmatrix}\begin{bmatrix}\beta_0' & \delta_0'\end{bmatrix}'\|_\infty \\
\leq &\sqrt{n}\max_{j\in H}\|\widehat{\Theta}_j(\hat{\tau})\|_1\sum_{j\in H}|g_j|\|\widehat{\boldsymbol{M}}(\tau_0)-\widehat{\boldsymbol{M}}(\hat{\tau})\|_\infty\|\delta_0\|_1
\end{aligned}$$

$$\begin{aligned}
&\|\widehat{\boldsymbol{M}}(\tau_0)-\widehat{\boldsymbol{M}}(\hat{\tau})\|_\infty \\
\leq &\max_{1\leq j,l\leq p}\left|\frac{1}{n}\sum_{i=1}^n X_i^{(j)}X_i^{(l)}[1(Q_i<\tau_0)-1(Q_i<\hat{\tau})]\right| \\
\leq &\max_{1\leq j,l\leq p}\sup_{|\tau-\tau_0|\leq|\tau_0-\widehat{\tau}|}\left|\frac{1}{n}\sum_{i=1}^n X_i^{(j)}X_i^{(l)}[1(Q_i<\tau_0)-1(Q_i<\tau)]\right| \\
\leq &C_5|\tau_0-\widehat{\tau}| = O_p\left(\frac{s_0\log p}{n}\right)
\end{aligned}$$

where the last equality follows from Assumption 4(1.3.5).

We know that $\sup_{\tau\in\mathbb{T}}\max_{j\in H}\|\widehat{\Theta}_j(\tau)\|_1 = O_p(\sqrt{\bar{s}})$ by Lemma 2

$$
\begin{aligned}
&|g'\hat{\Theta}(\hat{\tau})(\boldsymbol{X}'(\hat{\tau})\boldsymbol{X}(\tau_0)-\boldsymbol{X}'(\hat{\tau})\boldsymbol{X}(\hat{\tau}))\alpha_0/n^{1/2}| \\
\leq&\sqrt{n}\max_{j\in H}\|\widehat{\Theta}_j(\hat{\tau})\|_1\sum_{j\in H}|g_j|\|\widehat{\boldsymbol{M}}(\tau_0)-\widehat{\boldsymbol{M}}(\hat{\tau})\|_\infty\|\delta_0\|_1 \\
\leq&\sqrt{n}O_p(\sqrt{h\bar{s}})O_p\left(\frac{s_0\log p}{n}\right)\|\delta_0\|_1 \\
=&O_p\left(\frac{\|\delta_0\|_1 s_0\sqrt{h\bar{s}}\log p}{\sqrt{n}}\right)=O_p\left(\frac{s_0^2\sqrt{h\bar{s}}\log p}{\sqrt{n}}\right)
\end{aligned}
$$

□

Lemma 22. *Suppose that Assumption1 to 6 be satisfied, conditional on events* $\mathbb{A}_1$ *,*$\mathbb{A}_2$*,* $\mathbb{A}_3$*,* $\mathbb{A}_4$ *and* $\mathbb{A}_5$*, then*

$$
|g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g-g'\Theta(\hat{\tau})\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g|=O_p\left(h\bar{s}\sqrt{s_0^3}\sqrt{\frac{\log p}{n}}\right).
$$

Proof of Lemma 22 .

$$
\begin{aligned}
\text{Recall }\Sigma(\hat{\tau})_{xu}=&E\left[\boldsymbol{X}_i(\hat{\tau})\boldsymbol{X}_i'(\hat{\tau})U_i^2\right]=E\left[\boldsymbol{X}_i(\hat{\tau})\boldsymbol{X}_i'(\hat{\tau})\right]E\left[U_i^2\right], \\
\hat{U}_i(\hat{\tau})=&Y_i-\boldsymbol{X}_i'(\hat{\tau})\hat{\alpha}(\hat{\tau})=U_i+\boldsymbol{X}_i'(\tau_0)\alpha_0-\boldsymbol{X}_i'(\hat{\tau})\hat{\alpha}(\hat{\tau}), \\
\hat{\Sigma}(\hat{\tau})_{xu}=&\frac{1}{n}\sum_{i=1}^n\boldsymbol{X}_i(\hat{\tau})\boldsymbol{X}_i'(\hat{\tau})\hat{U}_i(\hat{\tau})^2, \\
\text{and set }\tilde{\Sigma}(\hat{\tau})_{xu}=&\frac{1}{n}\sum_{i=1}^n\boldsymbol{X}_i(\hat{\tau})\boldsymbol{X}_i'(\hat{\tau})U_i^2
\end{aligned}
$$

We first show that

$$
\begin{aligned}
&|g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g-g'\Theta(\hat{\tau})\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g| \\
=\;&|g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g-g'\widehat{\Theta}(\hat{\tau})\tilde{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g+g'\widehat{\Theta}(\hat{\tau})\tilde{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g-g'\Theta(\hat{\tau})\tilde{\Sigma}(\hat{\tau})_{xu}\Theta(\hat{\tau})'g \\
&+g'\Theta(\hat{\tau})\tilde{\Sigma}(\hat{\tau})_{xu}\Theta(\hat{\tau})'g-g'\Theta(\hat{\tau})\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g\|_\infty \\
\leq\;&|g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g-g'\widehat{\Theta}(\hat{\tau})\tilde{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g|+|g'\widehat{\Theta}(\hat{\tau})\tilde{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g-g'\Theta(\hat{\tau})\tilde{\Sigma}(\hat{\tau})_{xu}\Theta(\hat{\tau})'g\|_\infty \\
&+|g'\Theta(\hat{\tau})\tilde{\Sigma}(\hat{\tau})_{xu}\Theta(\hat{\tau})'g-g'\Theta(\hat{\tau})\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g|
\end{aligned}
$$

To prove this lemma,I need prove the followings

$$|g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g - g'\widehat{\Theta}(\hat{\tau})\tilde{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g| = o_p(1)$$

$$|g'\widehat{\Theta}(\hat{\tau})\tilde{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g - g'\widehat{\Theta}(\hat{\tau})\Sigma(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g| = o_p(1)$$

$$|g'\widehat{\Theta}(\hat{\tau})\Sigma(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g - g'\Theta(\hat{\tau})\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g| = o_p(1)$$

Step 1.

$$\begin{aligned} & |g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g - g'\widehat{\Theta}(\hat{\tau})\tilde{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g| \\ \leq \quad & |g'\widehat{\Theta}(\hat{\tau})\left(\hat{\Sigma}(\hat{\tau})_{xu} - \tilde{\Sigma}(\hat{\tau})_{xu}\right)\widehat{\Theta}(\hat{\tau})'g| \\ \leq \quad & ||g'\widehat{\Theta}(\hat{\tau})||_1^2||\hat{\Sigma}(\hat{\tau})_{xu} - \tilde{\Sigma}(\hat{\tau})_{xu}||_\infty \end{aligned}$$

Before I expand and simplify equations, I note that

$$\alpha_0'\boldsymbol{X}_i(\tau_0)\boldsymbol{X}_i'(\hat{\tau})\hat{\alpha}(\hat{\tau}) = \hat{\alpha}'(\hat{\tau})\boldsymbol{X}_i(\hat{\tau})\boldsymbol{X}_i'(\tau_0)\alpha_0$$

$$
\begin{aligned}
&\hat{\Sigma}(\hat{\tau})_{xu} - \tilde{\Sigma}(\hat{\tau})_{xu} \\
=&\frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{X}_i(\hat{\tau})\boldsymbol{X}_i'(\hat{\tau})\hat{U}_i^2(\hat{\tau}) - \boldsymbol{X}_i(\hat{\tau})\boldsymbol{X}_i'(\hat{\tau})U_i^2\right) \\
=&\frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{X}_i(\hat{\tau})\boldsymbol{X}_i'(\hat{\tau})(U_i + \boldsymbol{X}_i'(\tau_0)\alpha_0 - \boldsymbol{X}_i'(\hat{\tau})\hat{\alpha}(\hat{\tau}))^2 - \boldsymbol{X}_i(\hat{\tau})\boldsymbol{X}_i'(\hat{\tau})U_i^2\right) \\
=&\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i(\hat{\tau})\boldsymbol{X}_i'(\hat{\tau})U_i^2 \\
+&\frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{X}_i(\hat{\tau})\boldsymbol{X}_i'(\hat{\tau})\alpha_0'\boldsymbol{X}_i(\tau_0)\boldsymbol{X}_i'(\tau_0)\alpha_0\right) \\
+&\frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{X}_i(\hat{\tau})\boldsymbol{X}_i'(\hat{\tau})\hat{\alpha}'(\hat{\tau})\boldsymbol{X}_i(\hat{\tau})\boldsymbol{X}_i'(\hat{\tau})\hat{\alpha}(\hat{\tau})\right) \\
-&\frac{2}{n}\sum_{i=1}^{n}\left(\boldsymbol{X}_i(\hat{\tau})\boldsymbol{X}_i'(\hat{\tau})\alpha_0'\boldsymbol{X}_i(\tau_0)\boldsymbol{X}_i'(\hat{\tau})\hat{\alpha}(\hat{\tau})\right) \\
+&\frac{2}{n}\sum_{i=1}^{n}\left(\boldsymbol{X}_i(\hat{\tau})\boldsymbol{X}_i'(\hat{\tau})\alpha_0'\boldsymbol{X}_i(\tau_0)U_i\right) \\
-&\frac{2}{n}\sum_{i=1}^{n}\left(\boldsymbol{X}_i(\hat{\tau})\boldsymbol{X}_i'(\hat{\tau})\hat{\alpha}(\hat{\tau})'\boldsymbol{X}_i(\hat{\tau})U_i\right) \\
-&\frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{X}_i(\hat{\tau})\boldsymbol{X}_i'(\hat{\tau})U_i^2\right) \\
=&\frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{X}_i(\hat{\tau})\boldsymbol{X}_i'(\hat{\tau})\alpha_0\boldsymbol{X}_i'(\tau_0)\left(\boldsymbol{X}_i'(\tau_0)\alpha_0 - \boldsymbol{X}_i'(\hat{\tau})\hat{\alpha}(\hat{\tau})\right)\right) \\
+&\frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{X}_i(\hat{\tau})\boldsymbol{X}_i'(\hat{\tau})\hat{\alpha}(\hat{\tau})\boldsymbol{X}_i'(\hat{\tau})\left(\boldsymbol{X}_i'(\hat{\tau})\hat{\alpha}(\hat{\tau}) - \boldsymbol{X}_i'(\tau_0)\alpha_0\right)\right) \\
+&\frac{2}{n}\sum_{i=1}^{n}\left(\boldsymbol{X}_i(\hat{\tau})\boldsymbol{X}_i'(\hat{\tau})U_i\left(\alpha_0'\boldsymbol{X}_i(\tau_0) - \hat{\alpha}(\hat{\tau})'\boldsymbol{X}_i(\hat{\tau})\right)\right)
\end{aligned}
$$

Recall Lemma 16,

$$
\max_{1\le k,l,j\le p}\left|\frac{1}{n}\sum_{i=1}^{n}(X_i^{(k)}X_i^{(l)}X_i^{(j)})^2 - E\left[(X_i^{(k)}X_i^{(l)}X_i^{(j)})^2\right]\right| = O_p\left(\frac{\sqrt{\log p}}{\sqrt{n}}\right)
$$

$$
\max_{1\le k,l\le p}\left|\frac{1}{n}\sum_{i=1}^{n}(X_i^{(k)}X_i^{(l)}U_i)^2 - E\left[(X_i^{(k)}X_i^{(l)}U_i)^2\right]\right| = O_p\left(\frac{\sqrt{\log p}}{\sqrt{n}}\right)
$$

Applying Theorem 2,

$$\|\hat{\alpha}(\hat{\tau})\|_1 \leq \|\alpha_0\|_1 + O_p\left(s_0\sqrt{\frac{\log p}{n}}\right)$$

$$\|\boldsymbol{X}(\hat{\tau})\hat{\alpha}(\hat{\tau}) - \boldsymbol{X}(\tau_0)\alpha_0\|_n = O_p\left(\sqrt{s_0}\sqrt{\frac{\log p}{n}}\right)$$

$$\|\alpha_0\|_1 = O_p(s_0)$$

By Cauchy-Schwarz inequality and holder's inequality

$$\begin{aligned}
& \max_{1\leq k,l\leq 2p} \left|\tfrac{1}{n}\textstyle\sum_{i=1}^n \left(\boldsymbol{X}_i^{(k)}(\hat{\tau})\boldsymbol{X}_i^{(l)}(\hat{\tau})\alpha_0'\boldsymbol{X}_i(\tau_0)\left(\boldsymbol{X}_i'(\tau_0)\alpha_0 - \boldsymbol{X}_i'(\hat{\tau})\hat{\alpha}(\hat{\tau})\right)\right)\right| \\
\leq\ & \sqrt{\max_{1\leq k,l\leq 2p}\max_{1\leq i\leq n}\tfrac{1}{n}\textstyle\sum_{i=1}^n(\boldsymbol{X}_i^{(k)}(\hat{\tau})\boldsymbol{X}_i^{(l)}(\hat{\tau}))^2\left(\alpha_0\boldsymbol{X}_i'(\tau_0)\right)^2}\,\|\boldsymbol{X}(\tau_0)\alpha_0 - \boldsymbol{X}(\hat{\tau})\hat{\alpha}(\hat{\tau})\|_n \\
\leq\ & \sqrt{\max_{1\leq k,l\leq 2p}\max_{1\leq i\leq n}\tfrac{1}{n}\textstyle\sum_{i=1}^n(\boldsymbol{X}_i^{(k)}(\hat{\tau})\boldsymbol{X}_i^{(l)}(\hat{\tau}))^2\left(\max_{1\leq k\leq 2p}\boldsymbol{X}_i^{(k)}(\tau_0)\right)^2\|\alpha_0\|_1^2}\,\|\boldsymbol{X}(\tau_0)\alpha_0 - \boldsymbol{X}(\hat{\tau})\hat{\alpha}(\hat{\tau})\|_n \\
\leq\ & \sqrt{\max_{1\leq k,l,j\leq p}\tfrac{1}{n}\textstyle\sum_{i=1}^n(X_i^{(k)}X_i^{(l)}X_i^{(j)})^2\|\alpha_0\|_1^2\cdot 1(Q_i<\tau_0)\cdot 1(Q_i<\hat{\tau})}\,\|\boldsymbol{X}(\hat{\tau})\hat{\alpha}(\hat{\tau}) - \boldsymbol{X}(\tau_0)\alpha_0\|_n \\
\leq\ & O_p\left(\sqrt{s_0^3}\sqrt{\tfrac{\log p}{n}}\right)
\end{aligned}$$

$$\begin{aligned}
& \max_{1\leq k,l\leq 2p} \left|\tfrac{1}{n}\textstyle\sum_{i=1}^n \left(\boldsymbol{X}_i^{(k)}(\hat{\tau})\boldsymbol{X}_i^{(l)}(\hat{\tau})\hat{\alpha}(\hat{\tau})\boldsymbol{X}_i'(\hat{\tau})\left(\boldsymbol{X}_i'(\hat{\tau})\hat{\alpha}(\hat{\tau}) - \boldsymbol{X}_i'(\tau_0)\alpha_0\right)\right)\right. \\
\leq\ & \max_{1\leq k,l\leq 2p}\max_{1\leq i\leq n}\sqrt{\tfrac{1}{n}\textstyle\sum_{i=1}^n(\boldsymbol{X}_i^{(k)}(\hat{\tau})\boldsymbol{X}_i^{(l)}(\hat{\tau}))^2\left(\hat{\alpha}'(\hat{\tau})\boldsymbol{X}_i(\hat{\tau})\right)^2}\,\|\boldsymbol{X}(\hat{\tau})\hat{\alpha}(\hat{\tau}) - \boldsymbol{X}(\tau_0)\alpha_0\|_n \\
\leq\ & \sqrt{\max_{1\leq k,l,j\leq p}\tfrac{1}{n}\textstyle\sum_{i=1}^n(X_i^{(k)}X_i^{(l)}X_i^{(j)})^2\|\hat{\alpha}(\hat{\tau})\|_1^2|1(Q_i<\hat{\tau})}\,\|\boldsymbol{X}(\hat{\tau})\hat{\alpha}(\hat{\tau}) - \boldsymbol{X}(\tau_0)\alpha_0\|_n \\
\leq\ & O_p\left(\sqrt{s_0^3}\sqrt{\tfrac{\log p}{n}}\right)
\end{aligned}$$

$$\begin{aligned}
& \max_{1\leq k,l\leq 2p} \left|\tfrac{2}{n}\textstyle\sum_{i=1}^n \left(\boldsymbol{X}_i'(\tau_0)\alpha_0 - \boldsymbol{X}_i'(\hat{\tau})\hat{\alpha}(\hat{\tau})\right)(\boldsymbol{X}_i^{(k)}(\hat{\tau})\boldsymbol{X}_i^{(l)}(\hat{\tau}))U_i\right| \\
\leq\ & 2\sqrt{\max_{1\leq k,l\leq p}\tfrac{1}{n}\textstyle\sum_{i=1}^n(X_i^{(k)}X_i^{(l)}U_i)^2\cdot 1(Q_i<\hat{\tau})}\,\|\boldsymbol{X}(\tau_0)\alpha_0 - \boldsymbol{X}(\hat{\tau})\hat{\alpha}(\hat{\tau})\|_n \\
\leq\ & O_p\left(\sqrt{s_0}\sqrt{\tfrac{\log p}{n}}\right)
\end{aligned}$$

Hence,

$$\|\hat{\Sigma}(\hat{\tau})_{xu} - \tilde{\Sigma}(\hat{\tau})_{xu}\|_\infty = O_p\left(\sqrt{s_0^3}\sqrt{\frac{\log p}{n}}\right)$$

$$
\begin{aligned}
& |g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g - g'\widehat{\Theta}(\hat{\tau})\tilde{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g| \\
\leq \quad & \|g'\widehat{\Theta}(\hat{\tau})\|_1^2 \|\hat{\Sigma}(\hat{\tau})_{xu} - \tilde{\Sigma}(\hat{\tau})_{xu}\|_\infty \\
\leq \quad & \left(\sum_{j\in H}|g_j| \max_{j\in H}\sup_{\tau\in\mathbb{T}}\|\widehat{\Theta}(\hat{\tau})\|_1\right)^2 \|\hat{\Sigma}(\hat{\tau})_{xu} - \tilde{\Sigma}(\hat{\tau})_{xu}\|_\infty \\
= \quad & O_p(h\bar{s})\, O_p\left(\sqrt{s_0^3}\sqrt{\tfrac{\log p}{n}}\right) = O_p\left(h\bar{s}\sqrt{s_0^3}\sqrt{\tfrac{\log p}{n}}\right)
\end{aligned}
$$

Step 2.Next, I show that

$$
|g'\widehat{\Theta}(\hat{\tau})\tilde{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g - g'\widehat{\Theta}(\hat{\tau})\Sigma(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g| = o_p(1)
$$

Note that

$$
\begin{aligned}
& \tilde{\Sigma}(\hat{\tau})_{xu} - \Sigma(\hat{\tau})_{xu} \\
= \quad & \tfrac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i(\hat{\tau})\boldsymbol{X}_i'(\hat{\tau})U_i^2 - E\left[\boldsymbol{X}_i(\hat{\tau})\boldsymbol{X}_i'(\hat{\tau})U_i^2\right]
\end{aligned}
$$

Recall Lemma 16,

$$
\max_{1\leq l,k\leq 2p} \left|\frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i^{(k)}(\hat{\tau})\boldsymbol{X}_i^{(l)}(\hat{\tau})U_i^2 - E\left[\boldsymbol{X}_i^{(k)}(\hat{\tau})\boldsymbol{X}_i^{(l)}(\hat{\tau})U_i^2\right]\right| = O_p\left(\frac{\sqrt{\log p}}{\sqrt{n}}\right)
$$

Therefore

$$
\begin{aligned}
& |g'\widehat{\Theta}(\hat{\tau})\tilde{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g - g'\widehat{\Theta}(\hat{\tau})\Sigma(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g| \\
\leq \quad & |g'\widehat{\Theta}(\hat{\tau})\left(\tilde{\Sigma}(\hat{\tau})_{xu} - \Sigma(\hat{\tau})_{xu}\right)\widehat{\Theta}(\hat{\tau})'g| \\
\leq \quad & \|g'\widehat{\Theta}(\hat{\tau})\|_1^2 \|\tilde{\Sigma}(\hat{\tau})_{xu} - \Sigma(\hat{\tau})_{xu}\|_\infty \\
\leq \quad & \left(\sum_{j\in H}|g_j| \max_{j\in H}\sup_{\tau\in\mathbb{T}}\|\widehat{\Theta}_j(\hat{\tau})\|_1\right)^2 \|\tilde{\Sigma}(\hat{\tau})_{xu} - \Sigma(\hat{\tau})_{xu}\|_\infty \\
\leq \quad & O_p(h\bar{s})\, O_p\left(\sqrt{\tfrac{\log p}{n}}\right) = O_p\left(h\bar{s}\sqrt{\tfrac{\log p}{n}}\right)
\end{aligned}
$$

Step 3.Next, I show that

$$|g'\widehat{\Theta}(\hat{\tau})\Sigma(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g - g'\Theta(\hat{\tau})\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g| = o_p(1)$$

By Lemma 6.1 in van de Geer et al. (2014)

$$\begin{aligned}
& |g'\widehat{\Theta}(\hat{\tau})\Sigma(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g - g'\Theta(\hat{\tau})\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g| \\
\leq \; & \|\Sigma(\hat{\tau})_{xu}\|_\infty \|\left(\widehat{\Theta}(\hat{\tau}) - \Theta(\hat{\tau})\right)' g\|_1^2 + 2\|\left(\widehat{\Theta}(\hat{\tau}) - \Theta(\hat{\tau})\right)' g\|_2 \|\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g\|_2 \\
= \; & \|\Sigma(\hat{\tau})_{xu}\|_\infty \|\left(\widehat{\Theta}(\hat{\tau}) - \Theta(\hat{\tau})\right)' g\|_1^2 + 2\tilde{\kappa}(\bar{s}, c_0, \mathbb{T}, \Sigma_{xu})\|\left(\widehat{\Theta}(\hat{\tau}) - \Theta(\hat{\tau})\right)' g\|_2 \|\Theta(\hat{\tau})'g\|_2 \\
\leq \; & \|\Sigma(\hat{\tau})_{xu}\|_\infty \|\left(\widehat{\Theta}(\hat{\tau}) - \Theta(\hat{\tau})\right)' g\|_1^2 + 2\tilde{\kappa}(\bar{s}, c_0, \mathbb{T}, \Sigma_{xu})\|\left(\widehat{\Theta}(\hat{\tau}) - \Theta(\hat{\tau})\right)' g\|_2 \tilde{\kappa}(\bar{s}, c_0, \mathbb{T}, \Theta)\|g\|_2
\end{aligned}$$

As $\|\Sigma(\hat{\tau})_{xu}\|_\infty = \max_{1\leq l,k\leq 2p} E\left[\boldsymbol{X}_i^{(k)}(\hat{\tau})\boldsymbol{X}_i^{(l)}(\hat{\tau})u_i^2\right]$, $\tilde{\kappa}(\bar{s}, c_0, \mathbb{T}, \Sigma_{xu})$ and $\tilde{\kappa}(\bar{s}, c_0, \mathbb{T}, \Theta)$ are assumed bounded from Assumption 6,

$$\begin{aligned}
& \|\left(\widehat{\Theta}(\hat{\tau}) - \Theta(\hat{\tau})\right)' g\|_1 \\
= & \sum_{j\in H} \left(|g_j| \|\Theta_j(\hat{\tau}) - \Theta_j(\tau_0)\|_1\right) \\
\leq & \sum_{j\in H} |g_j| \sup_{\tau\in\widehat{\mathbb{T}}} \max_{j\in H} \|\Theta_j(\tau) - \Theta_j(\tau_0)\|_1 \\
\leq & \sqrt{h} \sup_{\tau\in\widehat{\mathbb{T}}} \max_{j\in H} \|\Theta_j(\tau) - \Theta_j(\tau_0)\|_1 \\
= & O_p\left(\sqrt{h}\bar{s}\sqrt{\frac{\log p}{n}}\right)
\end{aligned}$$

$$\begin{aligned}
& \|\left(\widehat{\Theta}(\hat{\tau}) - \Theta(\hat{\tau})\right)' g\|_2 \\
= & \|\sum_{j\in H} \left(\Theta_j(\hat{\tau}) - \Theta_j(\tau_0)\right)|g_j|\|_2 \\
\leq & \max_{j\in H} \|\Theta_j(\hat{\tau}) - \Theta_j(\tau_0)\|_2 \sum_{j\in H} |g_j| \\
\leq & \sqrt{h} \sup_{\tau\in\widehat{\mathbb{T}}} \max_{j\in H} \|\Theta_j(\tau) - \Theta_j(\tau_0)\|_2 \\
= & O_p\left(\sqrt{h\bar{s}}\sqrt{\frac{\log p}{n}}\right)
\end{aligned}$$

Furthermore,

$$
\begin{aligned}
& |g'\widehat{\Theta}(\hat\tau)\Sigma(\hat\tau)_{xu}\widehat{\Theta}(\hat\tau)'g - g'\Theta(\hat\tau)\Sigma(\hat\tau)_{xu}\Theta(\hat\tau)'g| \\
\leq \quad & \|\Sigma(\hat\tau)_{xu}\|_\infty \| \left(\widehat{\Theta}(\hat\tau) - \Theta(\hat\tau)\right)' g\|_1^2 + 2\tilde\kappa(\bar s, c_0, \mathbb{T}, \Sigma_{xu}) \| \left(\widehat{\Theta}(\hat\tau) - \Theta(\hat\tau)\right)' g\|_2 \tilde\kappa(\bar s, c_0, \mathbb{T}, \Theta)\|g\|_2 \\
\leq \quad & O_p\left(\sqrt{h}\bar s\sqrt{\tfrac{\log p}{n}}\right)^2 + O_p\left(\sqrt{h\bar s}\sqrt{\tfrac{\log p}{n}}\right) \\
= \quad & O_p\left(\sqrt{h\bar s}\sqrt{\tfrac{\log p}{n}}\right)
\end{aligned}
$$

Finally, by Assumption 6 (ii),

$$
\begin{aligned}
& |\widehat{\Theta}(\hat\tau)\hat\Sigma(\hat\tau)_{xu}\widehat{\Theta}(\hat\tau)' - \Theta(\hat\tau)\Sigma(\hat\tau)_{xu}\Theta(\hat\tau)'| \\
= \quad & O_p\left(h\bar s\sqrt{s_0^3}\sqrt{\tfrac{\log p}{n}}\right) + O_p\left(h\bar s\sqrt{\tfrac{\log p}{n}}\right) + O_p\left(\sqrt{h\bar s}\sqrt{\tfrac{\log p}{n}}\right) = O_p\left(h\bar s\sqrt{s_0^3}\sqrt{\tfrac{\log p}{n}}\right)
\end{aligned}
$$

□

Lemma 23. *Suppose that Assumption1 to 6 be satisfied, conditional on events* $\mathbb{A}_1$, $\mathbb{A}_2$, $\mathbb{A}_3$, $\mathbb{A}_4$, $\mathbb{A}_5$, *then*

$$|g'\widehat{\Theta}(\hat\tau)\widehat{\Sigma}(\hat\tau)_{xu}\widehat{\Theta}(\hat\tau)'g - g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g| = o_p(1).$$

Proof of Lemma 23.

$$
\begin{aligned}
& |g'\Theta(\hat\tau)\Sigma(\hat\tau)_{xu}\Theta(\hat\tau)'g - g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g| \\
=& |\Theta(\hat\tau)\Sigma(\hat\tau)_{xu}\Theta(\hat\tau)' - \Theta(\tau_0)\Sigma(\hat\tau)_{xu}\Theta(\hat\tau)' + g'\Theta(\tau_0)\Sigma(\hat\tau)_{xu}\Theta(\hat\tau)'g - g'\Theta(\tau_0)\Sigma(\hat\tau)_{xu}\Theta(\tau_0)'g \\
+& g'\Theta(\tau_0)\Sigma(\hat\tau)_{xu}\Theta(\tau_0)'g - g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g| \\
\leq& |g'\Theta(\hat\tau)\Sigma(\hat\tau)_{xu}\Theta(\hat\tau)'g - g'\Theta(\tau_0)\Sigma(\hat\tau)_{xu}\Theta(\hat\tau)'g| \\
+& |g'\Theta(\tau_0)\Sigma(\hat\tau)_{xu}\Theta(\hat\tau)'g - g'\Theta(\tau_0)\Sigma(\hat\tau)_{xu}\Theta(\tau_0)'g| \\
+& |g'\Theta(\tau_0)\Sigma(\hat\tau)_{xu}\Theta(\tau_0)'g - g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g|
\end{aligned}
$$

To prove this lemma, I need prove the followings

$$|g'\Theta(\hat\tau)\Sigma(\hat\tau)_{xu}\Theta(\hat\tau)'g - g'\Theta(\tau_0)\Sigma(\hat\tau)_{xu}\Theta(\hat\tau)'g| = o_p(1)$$

$$|g'\Theta(\tau_0)\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g - g'\Theta(\tau_0)\Sigma(\hat{\tau})_{xu}\Theta(\tau_0)'g| = o_p(1)$$

$$|g'\Theta(\tau_0)\Sigma(\hat{\tau})_{xu}\Theta(\tau_0)'g - g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g| = o_p(1)$$

Firstly, since $\Theta(\hat{\tau})$ is sysmetric, $\|\Theta(\hat{\tau})'g\|_1 = \|g'\Theta(\hat{\tau})\|_1$, also $\|\Sigma(\hat{\tau})_{xu}\|_\infty$ is bounded by Assumption 1, combine with (A.6.2)

$$\begin{aligned}
& |g'\Theta(\hat{\tau})\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g - g'\Theta(\tau_0)\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g| \\
\leq \quad & \|g'(\Theta(\hat{\tau}) - \Theta(\tau_0))\|_1 \|\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g\|_\infty \\
\leq \quad & \|g'(\Theta(\hat{\tau}) - \Theta(\tau_0))\|_1 \|\Sigma(\hat{\tau})_{xu}\|_\infty \|g'\Theta(\hat{\tau})\|_1 \\
\leq \quad & O_p\left(\sqrt{h}\bar{s}s_0\frac{\log p}{n}\right) O_p\left(\sqrt{h\bar{s}}\right) \\
= \quad & O_p\left(h\sqrt{\bar{s}^3}s_0\frac{\log p}{n}\right)
\end{aligned}$$

Secondly, as $\|g'(\Theta(\hat{\tau}) - \Theta(\tau_0))\|_1 = \|(\Theta(\hat{\tau}) - \Theta(\tau_0))'g\|_1$

$$\begin{aligned}
& |g'\Theta(\tau_0)\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g - g'\Theta(\tau_0)\Sigma(\hat{\tau})_{xu}\Theta(\tau_0)'g| \\
\leq \quad & \|g'\Theta(\hat{\tau})\|_1 \|\Sigma(\hat{\tau})_{xu}\|_\infty \|(\Theta(\hat{\tau}) - \Theta(\tau_0))'g\|_1 \\
= \quad & O_p\left(h\sqrt{\bar{s}^3}s_0\frac{\log p}{n}\right)
\end{aligned}$$

Note,

$$\begin{aligned}
& \Sigma(\hat{\tau})_{xu} - \Sigma(\tau_0)_{xu} \\
= \quad & E\left[\boldsymbol{X}_i(\hat{\tau})\boldsymbol{X}_i'(\hat{\tau})u_i^2\right] - E\left[\boldsymbol{X}_i(\tau_0)\boldsymbol{X}_i'(\tau_0)u_i^2\right] \\
= \quad & E\left[\boldsymbol{X}_i(\hat{\tau})\boldsymbol{X}_i'(\hat{\tau}) - \boldsymbol{X}_i(\tau_0)\boldsymbol{X}_i'(\tau_0)\right]E\left[u_i^2\right] \\
= \quad & (\Sigma(\hat{\tau}) - \Sigma(\tau_0))E\left[u_i^2\right] \\
= \quad & E\left[X_iX_i'\right]E\left[u_i^2\right]|\tau_0 - \widehat{\tau}|
\end{aligned}$$

and

$$\|\Sigma(\hat{\tau})_{xu}-\Sigma(\tau_0)_{xu}\|_\infty = \|\Sigma(\hat{\tau})-\Sigma(\tau_0)\|_\infty E\left[u_i^2\right] = O_p\left(s_0\frac{\log p}{n}\right)$$

$$\begin{aligned}
& |g'\Theta(\tau_0)\Sigma(\hat{\tau})_{xu}\Theta(\tau_0)'g - g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g| \\
\leq \quad & |g'\Theta(\tau_0)(\Sigma(\hat{\tau})_{xu}-\Sigma(\tau_0)_{xu})\Theta(\tau_0)'g| \\
\leq \quad & \|g'\Theta(\tau_0)\|_1^2\|\Sigma(\hat{\tau})_{xu}-\Sigma(\tau_0)_{xu}\|_\infty \\
= \quad & O_p(h\bar{s})\,O_p\left(s_0\tfrac{\log p}{n}\right) = O_p\left(h\bar{s}s_0\tfrac{\log p}{n}\right)
\end{aligned}$$

$$|g'\Theta(\hat{\tau})\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g - g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g| = O_p\left(h\sqrt{\bar{s}^3}s_0\frac{\log p}{n}\right) + O_p\left(h\bar{s}s_0\frac{\log p}{n}\right) = O_p\left(h\sqrt{\bar{s}^3}s_0\frac{\log p}{n}\right).$$

Combine with Lemma 22

$$|g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g - g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g| = O_p\left(h\bar{s}\sqrt{s_0^3}\sqrt{\frac{\log p}{n}}\right) + O_p\left(h\sqrt{\bar{s}^3}s_0\frac{\log p}{n}\right) = O_p\left(h\sqrt{s_0^3\bar{s}^3}\sqrt{\frac{\log p}{n}}\right)$$

□

Proof of Theorem 3 Case II: fixed threshold. We show that the ratio

(A.6.5)
$$t = \frac{\sqrt{n}g'(\hat{a}(\hat{\tau})-\alpha_0)}{\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}}$$

is asymptotically standard normal. First, note that by (1.4.11) one can write

$$t = t_1 + t_2,$$

where $t_1 = \frac{g'\widehat{\Theta}(\tau_0)\boldsymbol{X}'(\tau_0)U/n^{1/2}}{\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}}$, and

$$t_2 = \frac{g'(\hat{\Theta}(\hat{\tau})\boldsymbol{X}'(\hat{\tau})U - g'\hat{\Theta}(\tau_0)\boldsymbol{X}'(\tau_0)U)//n^{1/2} + g'\hat{\Theta}(\hat{\tau})(\boldsymbol{X}'(\hat{\tau})\boldsymbol{X}(\tau_0)-\boldsymbol{X}'(\hat{\tau})\boldsymbol{X}(\hat{\tau}))\alpha_0/n^{1/2} - g'\Delta(\hat{\tau})}{\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}}$$

It suffices to show that t_1 is asymptotically standard normal and $t_2 = o_p(1)$.

Step 1.

Step 1.1) This step, referring to the proof detailed in Step 1.1 for the proof of Theorem 3 for the no-threshold case, shows that in the fixed-threshold case, $t_1' = \frac{g'\Theta(\tau_0)\boldsymbol{X}'(\tau_0)U/n^{1/2}}{\sqrt{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)g}}$ is asymptotically standard normal.

Step 1.2). Let

$$t_1'' = \frac{g'\Theta(\tau_0)\boldsymbol{X}'(\tau_0)U/n^{1/2}}{\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})g}}$$

$$|g'\widehat{\Theta}(\tau_0)\boldsymbol{X}'(\tau_0)U/n^{1/2} - g'\Theta(\tau_0)\boldsymbol{X}'(\tau_0)U/n^{1/2}|$$

$$\leq \|g'\left(\widehat{\Theta}(\tau_0) - \Theta(\tau_0)\right)\|_1 \|\boldsymbol{X}(\tau_0)U/n^{1/2}\|_\infty$$

Conditional on $\mathbb{A}_1$, $\mathbb{A}_2$, $\mathbb{A}_3$ and $\mathbb{A}_4$ and by Lemma 2

$$= O_p(\sqrt{h}\bar{s}\frac{\sqrt{\log p}}{\sqrt{n}})O_p(\sqrt{\log p}) = O_p(\sqrt{h}\bar{s}\frac{\log p}{\sqrt{n}}) = o_p(1)$$

$$\begin{aligned}|t_1'' - t_1| &= \frac{g'\left(\widehat{\Theta}(\tau_0)\boldsymbol{X}'(\tau_0)U/n^{1/2} - \Theta(\tau_0)\boldsymbol{X}'(\tau_0)U/n^{1/2}\right)}{\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}} \\ &\leq o_p(1)\frac{1}{\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}}\end{aligned}$$

$$\begin{aligned}|t_1' - t_1''| &= \frac{\left(\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g} - \sqrt{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}\right)g'\Theta(\tau_0)\boldsymbol{X}'(\tau_0)u/n^{1/2}}{\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}\sqrt{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}} \\
&= \frac{\left(g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g - g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g\right)g'\Theta(\tau_0)\boldsymbol{X}'(\tau_0)u/n^{1/2}}{\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}\sqrt{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}\left(\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g} + \sqrt{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}\right)} \\
&\leq \frac{|g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g - g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g|g'\Theta(\tau_0)\boldsymbol{X}'(\tau_0)u/n^{1/2}}{\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}\sqrt{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}\left(\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g} + \sqrt{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}\right)} \\
&\leq \frac{O_p\left(h\sqrt{s_0^3\bar{s}^3}\sqrt{\frac{\log p}{n}}\right)O_p\left(\sqrt{h\bar{s}\log p}\right)}{\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}\sqrt{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}\left(\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g} + \sqrt{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}\right)} \\
&= \frac{O_p\left((hs_0)^{\frac{3}{2}}\bar{s}^2\frac{\log p}{\sqrt{n}}\right)}{\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}\sqrt{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}\left(\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g} + \sqrt{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}\right)} \\
&= \frac{o_p(1)}{\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}\sqrt{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}\left(\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g} + \sqrt{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}\right)}\end{aligned}$$

by Lemma 23.

Then combine the above two,

$$|t_1 - t_1'| \leq o_p(1)$$

Step 2. By Lemma 20, 21, and 18,

$$t_2 = \frac{g'(\hat{\Theta}(\hat{\tau})\boldsymbol{X}'(\hat{\tau})U - \hat{\Theta}(\tau_0)\boldsymbol{X}'(\tau_0)U)//n^{1/2} + g'\hat{\Theta}(\hat{\tau})(\boldsymbol{X}'(\hat{\tau})\boldsymbol{X}(\tau_0) - \boldsymbol{X}'(\hat{\tau})\boldsymbol{X}(\hat{\tau}))\alpha_0/n^{1/2} - g'\Delta(\hat{\tau})}{\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}} = o_p(1).$$

Step 3. Finally, by Slutsky's theorem

$$t = o_p(1) + t_1' \xrightarrow{d} N(0,1).$$

Therefore Lemma 23 implies that

$$\text{(A.6.6)} \qquad \sup_{\alpha_0 \in \mathscr{A}_{\ell_0}^{(2)}(s_0)} |\hat{\Theta}(\hat{\tau})\hat{\Sigma}_{xu}(\hat{\tau})\hat{\Theta}'(\hat{\tau}) - \Theta(\tau_0)\Sigma_{xu}(\tau_0)\Theta'(\tau_0)| = o_p(1)$$

where

$$\mathscr{A}_{\ell_0}^{(2)}(s_0) = \left\{\alpha_0 \in \mathbb{R}^{2p} \mid ||\alpha_0||_\infty \leq C, \mathscr{M}(\alpha_0) \leq s_0, \delta_0 \neq 0\right\}.$$

□

Proof of Theorem 4 . For $\varepsilon > 0$, define

$$\mathscr{F}_{1,n} = \left\{ \sup_{\alpha_0 \in \mathscr{B}_{\ell_0}(s_0)} |g'\Delta(\hat{\tau})| < \varepsilon \right\}$$

$$\mathscr{F}_{2,n} = \left\{ \sup_{\alpha_0 \in \mathscr{B}_{\ell_0}(s_0)} \left| \frac{g'\widehat{\Theta}(\widehat{\tau})\hat{\Sigma}(\widehat{\tau})_{xu}\widehat{\Theta}(\widehat{\tau})'g}{g'\Theta(\widehat{\tau})\Sigma(\widehat{\tau})_{xu}\Theta(\widehat{\tau})'g} - 1 \right| < \varepsilon \right\}$$

$$\mathscr{F}_{3,n} = \left\{ \sup_{\alpha_0 \in \mathscr{B}_{\ell_0}(s_0)} |g'\widehat{\Theta}(\widehat{\tau})X'(\widehat{\tau})U/n^{1/2} - g'\Theta(\widehat{\tau})X'(\widehat{\tau})U/n^{1/2}| < \varepsilon \right\}$$

$$\mathscr{F}_{4,n} = \left\{ \sup_{\alpha_0 \in \mathscr{A}_{\ell_0}^{(2)}(s_0)} |g'\hat{\Theta}(\hat{\tau})(\boldsymbol{X}'(\hat{\tau})\boldsymbol{X}(\tau_0) - \boldsymbol{X}'(\hat{\tau})\boldsymbol{X}(\hat{\tau}))\alpha_0/n^{1/2}| < \varepsilon \right\}$$

$$\mathscr{F}_{5,n} = \left\{ \sup_{\alpha_0 \in \mathscr{A}^{(2)}_{\ell_0}(s_0)} |g'(\hat{\Theta}(\hat{\tau})\boldsymbol{X}'(\hat{\tau})U - \hat{\Theta}(\tau_0)\boldsymbol{X}'(\tau_0)U)//n^{1/2}| < \varepsilon \right\}$$

$$\mathscr{F}_{6,n} = \left\{ \sup_{\alpha_0 \in \mathscr{A}^{(2)}_{\ell_0}(s_0)} |g'\widehat{\Theta}(\tau_0)X'(\tau_0)U/n^{1/2} - g'\Theta(\tau_0)X'(\tau_0)U/n^{1/2}| < \varepsilon \right\}$$

$$\mathscr{F}_{7,n} = \left\{ \sup_{\alpha_0 \in \mathscr{A}^{(2)}_{\ell_0}(s_0)} \left| \frac{g'\widehat{\Theta}(\widehat{\tau})\hat{\Sigma}(\widehat{\tau})_{xu}\widehat{\Theta}(\widehat{\tau})'g}{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g} - 1 \right| < \varepsilon \right\}$$

Applying Lemma 15 (and 18), 17, 2,21, 20, 19, and 23, respectively, I observe that the probabilities of these sets all approach one. Thus for every $t \in \mathbb{R}$,

(A.6.7)

$$\begin{aligned}
&\left| \mathbb{P}\left\{ \frac{\sqrt{n}g'(\hat{a}(\widehat{\tau}) - \alpha_0)}{\sqrt{g'\widehat{\Theta}(\widehat{\tau})\hat{\Sigma}(\widehat{\tau})_{xu}\widehat{\Theta}(\widehat{\tau})'g}} \le t \right\} - \Phi(t) \right| \\
\le &\left| \mathbb{P}\{\delta_0 \neq 0\} \mathbb{P}\left\{ \frac{g'\hat{\Theta}(\tau_0)\boldsymbol{X}'(\tau_0)U//n^{1/2} - g'\Delta(\hat{\tau}) + g'(\hat{\Theta}(\hat{\tau})\boldsymbol{X}'(\hat{\tau})U - \hat{\Theta}(\tau_0)\boldsymbol{X}'(\tau_0)U)/n^{1/2} + g'\hat{\Theta}(\hat{\tau})(\boldsymbol{X}'(\hat{\tau})\boldsymbol{X}(\tau_0) - \boldsymbol{X}'(\hat{\tau})\boldsymbol{X}(\hat{\tau}))\alpha_0/n^{1/2}}{\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}} \le t \right\} \right. \\
&\left. + \mathbb{P}\{\delta_0 = 0\} \mathbb{P}\left\{ \frac{g'\hat{\Theta}(\widehat{\tau})\boldsymbol{X}'(\widehat{\tau})U/n^{1/2} - g'\Delta(\hat{\tau})}{\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}} \le t \right\} - \Phi(t) \right| \\
\le &\mathbb{P}\{\delta_0 \neq 0\} \left| \mathbb{P}\left\{ \frac{g'\hat{\Theta}(\tau_0)\boldsymbol{X}'(\tau_0)U//n^{1/2} - g'\Delta(\hat{\tau}) + g'(\hat{\Theta}(\hat{\tau})\boldsymbol{X}'(\hat{\tau})U - \hat{\Theta}(\tau_0)\boldsymbol{X}'(\tau_0)U)/n^{1/2} + g'\hat{\Theta}(\hat{\tau})(\boldsymbol{X}'(\hat{\tau})\boldsymbol{X}(\tau_0) - \boldsymbol{X}'(\hat{\tau})\boldsymbol{X}(\hat{\tau}))\alpha_0/n^{1/2}}{\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}} \le t \right\} - \Phi(t) \right| \\
&+ \mathbb{P}\{\delta_0 = 0\} \left| \mathbb{P}\left\{ \frac{g'\hat{\Theta}(\widehat{\tau})\boldsymbol{X}'(\widehat{\tau})U/n^{1/2} - g'\Delta(\hat{\tau})}{\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}} \le t \right\} - \Phi(t) \right|
\end{aligned}$$

where $\mathbb{P}\{\delta_0 = 0\} + \mathbb{P}\{\delta_0 \neq 0\} = 1$, and these probabilities are between 0 and 1. Let's first consider the term in the final inequality of (A.6.7) for the case without a threshold

(A.6.8)

$$\begin{aligned}
&\left| \mathbb{P}\left\{ \frac{g'\hat{\Theta}(\widehat{\tau})\boldsymbol{X}'(\widehat{\tau})U/n^{1/2} - g'\Delta(\hat{\tau})}{\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}} \le t \right\} - \Phi(t) \right| \\
\le &\mathbb{P}\left\{ \frac{g'\hat{\Theta}(\widehat{\tau})\boldsymbol{X}'(\widehat{\tau})U/n^{1/2} - g'\Delta(\hat{\tau})}{\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}} \le t, \mathscr{F}_{1,n}, \mathscr{F}_{2,n}, \mathscr{F}_{3,n} \right\} + \mathbb{P}\left\{ \mathscr{F}^c_{1,n} \cup \mathscr{F}^c_{2,n} \cup \mathscr{F}^c_{3,n} \right\}.
\end{aligned}$$

There exists a positive constant D_1 such that

$$
\begin{aligned}
&\mathbb{P}\left\{\frac{g'\hat{\Theta}(\hat{\tau})\boldsymbol{X}'(\hat{\tau})U/n^{1/2}-g'\Delta(\hat{\tau})}{\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}}\le t,\mathscr{F}_{1,n},\mathscr{F}_{2,n},\mathscr{F}_{3,n}\right\}\\
=&\mathbb{P}\left\{\frac{g'\hat{\Theta}(\hat{\tau})\boldsymbol{X}'(\hat{\tau})U/n^{1/2}-g'\Delta(\hat{\tau})}{\sqrt{g'\Theta(\hat{\tau})\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g}}\le t\sqrt{\frac{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}{g'\Theta(\hat{\tau})\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g}},\mathscr{F}_{1,n},\mathscr{F}_{2,n},\mathscr{F}_{3,n}\right\}\\
\le&\mathbb{P}\left\{\frac{g'\Theta(\hat{\tau})X'(\hat{\tau})U/n^{1/2}}{\sqrt{g'\Theta(\hat{\tau})\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g}}\le t(1+\varepsilon)+\frac{\varepsilon+\varepsilon}{\sqrt{g'\Theta(\hat{\tau})\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g}}\right\}\\
\le&\mathbb{P}\left\{\frac{g'\Theta(\hat{\tau})X'(\hat{\tau})U/n^{1/2}}{\sqrt{g'\Theta(\hat{\tau})\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g}}\le t(1+\varepsilon)+D_1\varepsilon\right\}\\
\le&\Phi(t(1+\varepsilon)+D_1\varepsilon)+\varepsilon
\end{aligned}
\tag{A.6.9}
$$

where the last inequality is derived from the proof of Theorem 3, where I established the asymptotic normality of $\frac{g'\Theta(\hat{\tau})X'(\hat{\tau})U/n^{1/2}}{\sqrt{g'\Theta(\hat{\tau})\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g}}$. Therefore, since the right-hand sides in (A.6.9) do not depend on α_0, I obtain

$$
\sup_{\alpha_0\in\mathscr{A}^{(1)}_{\ell_0}(s_0)}\mathbb{P}\left\{\frac{g'\hat{\Theta}(\hat{\tau})\boldsymbol{X}'(\hat{\tau})U/n^{1/2}-g'\Delta(\hat{\tau})}{\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}}\le t,\mathscr{F}_{1,n},\mathscr{F}_{2,n},\mathscr{F}_{3,n}\right\}\le\mathbb{P}\left\{\frac{g'\Theta(\hat{\tau})X'(\hat{\tau})U/n^{1/2}}{\sqrt{g'\Theta(\hat{\tau})\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g}}\le t(1+\varepsilon)+D_1\varepsilon\right\}.
\tag{A.6.10}
$$

The above arguments hold for all $\varepsilon>0$. By the continuity of $\Phi(\cdot)$, for any $\eta>0$, I can choose ε to be sufficiently small and conclude that

$$
\sup_{\alpha_0\in\mathscr{A}^{(1)}_{\ell_0}(s_0)}\mathbb{P}\left\{\frac{g'\hat{\Theta}(\hat{\tau})\boldsymbol{X}'(\hat{\tau})U/n^{1/2}-g'\Delta(\hat{\tau})}{\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}}\le t,\mathscr{F}_{1,n},\mathscr{F}_{2,n},\mathscr{F}_{3,n}\right\}\le\Phi(t)+\eta+\varepsilon,
\tag{A.6.11}
$$

Next, considering that $g'\Theta(\hat{\tau})\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g$ is bounded away from zero, there exists

$$
\begin{aligned}
&\mathbb{P}\left\{\frac{g'\hat{\Theta}(\hat{\tau})\boldsymbol{X}'(\hat{\tau})U/n^{1/2}-g'\Delta(\hat{\tau})}{\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}}\le t,\mathscr{F}_{1,n},\mathscr{F}_{2,n},\mathscr{F}_{3,n}\right\}\\
=&\mathbb{P}\left\{\frac{g'\hat{\Theta}(\hat{\tau})\boldsymbol{X}'(\hat{\tau})U/n^{1/2}-g'\Delta(\hat{\tau})}{\sqrt{g'\Theta(\hat{\tau})\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g}}\le t\sqrt{\frac{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}{g'\Theta(\hat{\tau})\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g}},\mathscr{F}_{1,n},\mathscr{F}_{2,n},\mathscr{F}_{3,n}\right\}\\
\ge&\mathbb{P}\left\{\frac{g'\Theta(\hat{\tau})X'(\hat{\tau})U/n^{1/2}}{\sqrt{g'\Theta(\hat{\tau})\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g}}\le t(1-\varepsilon)-\frac{\varepsilon+\varepsilon}{\sqrt{g'\Theta(\hat{\tau})\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g}},\mathscr{F}_{1,n},\mathscr{F}_{2,n},\mathscr{F}_{3,n}\right\}\\
\ge&\mathbb{P}\left\{\frac{g'\Theta(\hat{\tau})\boldsymbol{X}'(\hat{\tau})U/n^{1/2}}{\sqrt{g'\Theta(\hat{\tau})\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g}}\le t(1-\varepsilon)-D_1\varepsilon\right\}+\mathbb{P}\left\{\mathscr{F}_{1,n}\cap\mathscr{F}_{2,n}\cap\mathscr{F}_{3,n}\right\}-1\\
\ge&\Phi(t(1-\varepsilon)-D_1\varepsilon)-\varepsilon+\mathbb{P}\left\{\mathscr{F}_{1,n}\cap\mathscr{F}_{2,n}\cap\mathscr{F}_{3,n}\right\}-1,
\end{aligned}
\tag{A.6.12}
$$

where the last inequality arises from the asymptotic normality of $\frac{g'\Theta(\hat{\tau})\boldsymbol{X}'(\hat{\tau})U/n^{1/2}}{\sqrt{g'\Theta(\hat{\tau})\Sigma(\hat{\tau})_{xu}\Theta(\hat{\tau})'g}}$.

As $\mathbb{P}\left\{\mathscr{F}_{1,n}\cap\mathscr{F}_{2,n}\cap\mathscr{F}_{3,n}\right\}$ can be made arbitrarily close to one by choosing n sufficiently large and ε sufficiently small, I have

$$
\inf_{\alpha_0\in\mathscr{A}^{(1)}_{\ell_0}(s_0)}\mathbb{P}\left\{\frac{g'\hat{\Theta}(\hat{\tau})\boldsymbol{X}'(\hat{\tau})U/n^{1/2}-g'\Delta(\hat{\tau})}{\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}}\le t,\mathscr{F}_{1,n},\mathscr{F}_{2,n},\mathscr{F}_{3,n}\right\}\ge\Phi(t(1-\varepsilon)-D_1\varepsilon)-\varepsilon.
\tag{A.6.13}
$$

By the continuity of $\Phi(\cdot)$, for any $\eta>0$, I can choose ε to be sufficiently small and conclude that

$$
\inf_{\alpha_0\in\mathscr{A}^{(1)}_{\ell_0}(s_0)}\mathbb{P}\left\{\frac{g'\hat{\Theta}(\hat{\tau})\boldsymbol{X}'(\hat{\tau})U/n^{1/2}-g'\Delta(\hat{\tau})}{\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}}\le t,\mathscr{F}_{1,n},\mathscr{F}_{2,n},\mathscr{F}_{3,n}\right\}\ge\Phi(t)-\eta-2\varepsilon
\tag{A.6.14}
$$

Combining (A.6.11) and (A.6.14), since $\sup_{\alpha_0\in\mathscr{A}^{(1)}_{\ell_0}(s_0)}\mathbb{P}\left\{\mathscr{F}^c_{1,n}\cup\mathscr{F}^c_{2,n}\cup\mathscr{F}^c_{3,n}\right\}\to 0$, I obtain

$$
\left|\sup_{\alpha_0\in\mathscr{A}^{(1)}_{\ell_0}(s_0)}\mathbb{P}\left\{\frac{\sqrt{n}g'(\hat{a}(\hat{\tau})-\alpha_0)}{\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}}\le t\right\}-\Phi(t)\right|\to 0.
\tag{A.6.15}
$$

Considering the term in the final inequality of (A.6.7) for the case with a fixed threshold

(A.6.16)

$$\left|\mathbb{P}\left\{\frac{g'\hat{\Theta}(\tau_0)\boldsymbol{X}'(\tau_0)U//n^{1/2}-g'\Delta(\hat{\tau})+g'(\hat{\Theta}(\hat{\tau})\boldsymbol{X}'(\hat{\tau})U-\hat{\Theta}(\tau_0)\boldsymbol{X}'(\tau_0)U)/n^{1/2}+g'\hat{\Theta}(\hat{\tau})(\boldsymbol{X}'(\hat{\tau})\boldsymbol{X}(\tau_0)-\boldsymbol{X}'(\hat{\tau})\boldsymbol{X}(\hat{\tau}))\alpha_0/n^{1/2}}{\sqrt{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}}\leq t\right\}-\Phi(t)\right|$$

$$\leq\mathbb{P}\left\{\frac{g'\hat{\Theta}(\tau_0)\boldsymbol{X}'(\tau_0)U//n^{1/2}-g'\Delta(\hat{\tau})+g'(\hat{\Theta}(\hat{\tau})\boldsymbol{X}'(\hat{\tau})U-\hat{\Theta}(\tau_0)\boldsymbol{X}'(\tau_0)U)/n^{1/2}+g'\hat{\Theta}(\hat{\tau})(\boldsymbol{X}'(\hat{\tau})\boldsymbol{X}(\tau_0)-\boldsymbol{X}'(\hat{\tau})\boldsymbol{X}(\hat{\tau}))\alpha_0/n^{1/2}}{\sqrt{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}}\leq t\sqrt{\frac{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}},\mathscr{F}_{1,n},\mathscr{F}_{4,n},\mathscr{F}_{5,n},\mathscr{F}_{6,n},\mathscr{F}_{7,n}\right\}$$

$$+\mathbb{P}\left\{\mathscr{F}_{1,n}^c\cup\mathscr{F}_{4,n}^c\cup\mathscr{F}_{5,n}^c\cup\mathscr{F}_{6,n}^c\cup\mathscr{F}_{7,n}^c\right\}$$

Since $g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g$ does not depend on α_0 and is bounded away from zero, there exists a positive constant D_2 such that

(A.6.17)

$$\mathbb{P}\left\{\frac{g'\hat{\Theta}(\tau_0)\boldsymbol{X}'(\tau_0)U//n^{1/2}-g'\Delta(\hat{\tau})+g'(\hat{\Theta}(\hat{\tau})\boldsymbol{X}'(\hat{\tau})U-\hat{\Theta}(\tau_0)\boldsymbol{X}'(\tau_0)U)/n^{1/2}+g'\hat{\Theta}(\hat{\tau})(\boldsymbol{X}'(\hat{\tau})\boldsymbol{X}(\tau_0)-\boldsymbol{X}'(\hat{\tau})\boldsymbol{X}(\hat{\tau}))\alpha_0/n^{1/2}}{\sqrt{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}}\leq t\sqrt{\frac{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}},\mathscr{F}_{1,n},\mathscr{F}_{4,n},\mathscr{F}_{5,n},\mathscr{F}_{6,n},\mathscr{F}_{7,n}\right\}$$

$$\leq\mathbb{P}\left\{\frac{g'\Theta(\tau_0)\boldsymbol{X}'(\tau_0)U//n^{1/2}}{\sqrt{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}}\leq t(1+\varepsilon)+\frac{4\varepsilon}{\sqrt{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}}\right\}$$

$$\leq\mathbb{P}\left\{\frac{g'\hat{\Theta}(\tau_0)\boldsymbol{X}'(\tau_0)U//n^{1/2}}{\sqrt{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}}\leq t(1+\varepsilon)+D_2\varepsilon\right\}$$

Therefore, since the right-hand sides in (A.6.17) do not depend on α_0, I obtain

(A.6.18)

$$\sup_{\alpha_0\in\mathscr{A}_{\ell_0}^{(2)}(s_0)}\mathbb{P}\left\{\frac{g'\hat{\Theta}(\tau_0)\boldsymbol{X}'(\tau_0)U//n^{1/2}-g'\Delta(\hat{\tau})+g'(\hat{\Theta}(\hat{\tau})\boldsymbol{X}'(\hat{\tau})U-\hat{\Theta}(\tau_0)\boldsymbol{X}'(\tau_0)U)/n^{1/2}+g'\hat{\Theta}(\hat{\tau})(\boldsymbol{X}'(\hat{\tau})\boldsymbol{X}(\tau_0)-\boldsymbol{X}'(\hat{\tau})\boldsymbol{X}(\hat{\tau}))\alpha_0/n^{1/2}}{\sqrt{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}}\leq t\sqrt{\frac{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}},\mathscr{F}_{1,n},\mathscr{F}_{4,n},\mathscr{F}_{5,n},\mathscr{F}_{6,n},\mathscr{F}_{7,n}\right\}$$

$$\leq\mathbb{P}\left\{\frac{g'\hat{\Theta}(\tau_0)\boldsymbol{X}'(\tau_0)U//n^{1/2}}{\sqrt{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}}\leq t(1+\varepsilon)+D_2\varepsilon\right\}$$

In the proof of Theorem 3, I established the asymptotic normality of $\frac{g'\hat{\Theta}(\tau_0)\boldsymbol{X}'(\tau_0)U//n^{1/2}}{\sqrt{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}}$. Then, for sufficiently large n,

(A.6.19)

$$\sup_{\alpha_0\in\mathscr{A}_{\ell_0}^{(2)}(s_0)}\mathbb{P}\left\{\frac{g'\hat{\Theta}(\tau_0)\boldsymbol{X}'(\tau_0)U//n^{1/2}-g'\Delta(\hat{\tau})+g'(\hat{\Theta}(\hat{\tau})\boldsymbol{X}'(\hat{\tau})U-\hat{\Theta}(\tau_0)\boldsymbol{X}'(\tau_0)U)/n^{1/2}+g'\hat{\Theta}(\hat{\tau})(\boldsymbol{X}'(\hat{\tau})\boldsymbol{X}(\tau_0)-\boldsymbol{X}'(\hat{\tau})\boldsymbol{X}(\hat{\tau}))\alpha_0/n^{1/2}}{\sqrt{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}}\leq t\sqrt{\frac{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}},\mathscr{F}_{1,n},\mathscr{F}_{4,n},\mathscr{F}_{5,n},\mathscr{F}_{6,n},\mathscr{F}_{7,n}\right\}$$

$$\leq\Phi(t(1+\varepsilon)+D_2\varepsilon)+\varepsilon.$$

The above arguments hold for all $\varepsilon>0$. By the continuity of $\Phi(\cdot)$, for any $\eta>0$, I can choose ε to be sufficiently small and conclude that

(A.6.20)

$$\sup_{\alpha_0\in\mathscr{A}_{\ell_0}^{(2)}(s_0)}\mathbb{P}\left\{\frac{g'\hat{\Theta}(\tau_0)\boldsymbol{X}'(\tau_0)U//n^{1/2}-g'\Delta(\hat{\tau})+g'(\hat{\Theta}(\hat{\tau})\boldsymbol{X}'(\hat{\tau})U-\hat{\Theta}(\tau_0)\boldsymbol{X}'(\tau_0)U)/n^{1/2}+g'\hat{\Theta}(\hat{\tau})(\boldsymbol{X}'(\hat{\tau})\boldsymbol{X}(\tau_0)-\boldsymbol{X}'(\hat{\tau})\boldsymbol{X}(\hat{\tau}))\alpha_0/n^{1/2}}{\sqrt{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}}\leq t\sqrt{\frac{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}},\mathscr{F}_{1,n},\mathscr{F}_{4,n},\mathscr{F}_{5,n},\mathscr{F}_{6,n},\mathscr{F}_{7,n}\right\}$$

$$\leq\Phi(t)+\eta+\varepsilon.$$

Next, considering that $g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g$ does not depend on α_0 and is bounded away from zero, there exists

(A.6.21)

$$\mathbb{P}\left\{\frac{g'\hat{\Theta}(\tau_0)\boldsymbol{X}'(\tau_0)U//n^{1/2}-g'\Delta(\hat{\tau})+g'(\hat{\Theta}(\hat{\tau})\boldsymbol{X}'(\hat{\tau})U-\hat{\Theta}(\tau_0)\boldsymbol{X}'(\tau_0)U)/n^{1/2}+g'\hat{\Theta}(\hat{\tau})(\boldsymbol{X}'(\hat{\tau})\boldsymbol{X}(\tau_0)-\boldsymbol{X}'(\hat{\tau})\boldsymbol{X}(\hat{\tau}))\alpha_0/n^{1/2}}{\sqrt{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}}\leq t\sqrt{\frac{g'\widehat{\Theta}(\hat{\tau})\hat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g}{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}},\mathscr{F}_{1,n},\mathscr{F}_{4,n},\mathscr{F}_{5,n},\mathscr{F}_{6,n},\mathscr{F}_{7,n}\right\}$$

$$\geq\mathbb{P}\left\{\frac{g'\Theta(\tau_0)\boldsymbol{X}'(\tau_0)U//n^{1/2}}{\sqrt{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}}\leq t(1+\varepsilon)-\frac{4\varepsilon}{\sqrt{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}},\mathscr{F}_{1,n},\mathscr{F}_{4,n},\mathscr{F}_{5,n},\mathscr{F}_{6,n},\mathscr{F}_{7,n}\right\}$$

$$\geq\mathbb{P}\left\{\frac{g'\Theta(\tau_0)\boldsymbol{X}'(\tau_0)U//n^{1/2}}{\sqrt{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}}\leq t(1+\varepsilon)-D_2\varepsilon\right\}+\mathbb{P}\left\{\mathscr{F}_{1,n}\cap\mathscr{F}_{4,n}\cap\mathscr{F}_{5,n}\cap\mathscr{F}_{6,n}\cap\mathscr{F}_{7,n}\right\}-1.$$

As the right-hand sides in the above display do not depend on α_0, and

$\mathbb{P}\left\{\mathscr{F}_{1,n}\cap\mathscr{F}_{4,n}\cap\mathscr{F}_{5,n}\cap\mathscr{F}_{6,n}\cap\mathscr{F}_{7,n}\right\}$ can be made arbitrarily close to one by choosing n sufficiently large and ε sufficiently small, I have

(A.6.22)

$$\inf_{\alpha_0 \in \mathscr{A}^{(2)}_{\ell_0}(s_0)} \mathbb{P}\left\{ \frac{g'\hat{\Theta}(\tau_0)\boldsymbol{X}'(\tau_0)U//n^{1/2} - g'\Delta(\hat{\tau}) + g'(\hat{\Theta}(\hat{\tau})\boldsymbol{X}'(\hat{\tau})U - \hat{\Theta}(\tau_0)\boldsymbol{X}'(\tau_0)U)/n^{1/2} + g'\hat{\Theta}(\hat{\tau})(\boldsymbol{X}'(\hat{\tau})\boldsymbol{X}(\tau_0) - \boldsymbol{X}'(\hat{\tau})\boldsymbol{X}(\hat{\tau}))\alpha_0/n^{1/2}}{\sqrt{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}} \leq t\sqrt{\frac{g'\widehat{\Theta}(\widehat{\tau})\hat{\Sigma}(\widehat{\tau})_{xu}\widehat{\Theta}(\widehat{\tau})'g}{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}}, \mathscr{F}_{1,n}, \mathscr{F}_{4,n}, \mathscr{F}_{5,n}, \mathscr{F}_{6,n}, \mathscr{F}_{7,n} \right\}$$
$$\geq \mathbb{P}\left\{ \frac{g'\hat{\Theta}(\tau_0)\boldsymbol{X}'(\tau_0)U//n^{1/2}}{\sqrt{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}} \leq t(1-\varepsilon) - D_2\varepsilon \right\} - \varepsilon.$$

In the proof of Theorem 3, I established the asymptotic normality of $\frac{g'\hat{\Theta}(\tau_0)\boldsymbol{X}'(\tau_0)U//n^{1/2}}{\sqrt{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}}$.Then, for sufficiently large n,

(A.6.23)

$$\inf_{\alpha_0 \in \mathscr{A}^{(2)}_{\ell_0}(s_0)} \mathbb{P}\left\{ \frac{g'\hat{\Theta}(\tau_0)\boldsymbol{X}'(\tau_0)U//n^{1/2} - g'\Delta(\hat{\tau}) + g'(\hat{\Theta}(\hat{\tau})\boldsymbol{X}'(\hat{\tau})U - \hat{\Theta}(\tau_0)\boldsymbol{X}'(\tau_0)U)/n^{1/2} + g'\hat{\Theta}(\hat{\tau})(\boldsymbol{X}'(\hat{\tau})\boldsymbol{X}(\tau_0) - \boldsymbol{X}'(\hat{\tau})\boldsymbol{X}(\hat{\tau}))\alpha_0/n^{1/2}}{\sqrt{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}} \leq t\sqrt{\frac{g'\widehat{\Theta}(\widehat{\tau})\hat{\Sigma}(\widehat{\tau})_{xu}\widehat{\Theta}(\widehat{\tau})'g}{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}}, \mathscr{F}_{1,n}, \mathscr{F}_{4,n}, \mathscr{F}_{5,n}, \mathscr{F}_{6,n}, \mathscr{F}_{7,n} \right\}$$
$$\geq \Phi(t(1-\varepsilon) - D_2\varepsilon) - \varepsilon$$

By the continuity of $\Phi(\cdot)$, for any $\eta > 0$, I can choose ε to be sufficiently small and conclude that

(A.6.24)

$$\inf_{\alpha_0 \in \mathscr{A}^{(2)}_{\ell_0}(s_0)} \mathbb{P}\left\{ \frac{g'\hat{\Theta}(\tau_0)\boldsymbol{X}'(\tau_0)U//n^{1/2} - g'\Delta(\hat{\tau}) + g'(\hat{\Theta}(\hat{\tau})\boldsymbol{X}'(\hat{\tau})U - \hat{\Theta}(\tau_0)\boldsymbol{X}'(\tau_0)U)/n^{1/2} + g'\hat{\Theta}(\hat{\tau})(\boldsymbol{X}'(\hat{\tau})\boldsymbol{X}(\tau_0) - \boldsymbol{X}'(\hat{\tau})\boldsymbol{X}(\hat{\tau}))\alpha_0/n^{1/2}}{\sqrt{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}} \leq t\sqrt{\frac{g'\widehat{\Theta}(\widehat{\tau})\hat{\Sigma}(\widehat{\tau})_{xu}\widehat{\Theta}(\widehat{\tau})'g}{g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g}}, \mathscr{F}_{1,n}, \mathscr{F}_{4,n}, \mathscr{F}_{5,n}, \mathscr{F}_{6,n}, \mathscr{F}_{7,n} \right\}$$
$$\geq \Phi(t) - \eta - 2\varepsilon$$

Combining (A.6.20) and (A.6.24), since $\sup_{\alpha_0 \in \mathscr{A}^{(2)}_{\ell_0}(s_0)} \mathbb{P}\left\{\mathscr{F}^c_{1,n} \cup \mathscr{F}^c_{4,n} \cup \mathscr{F}^c_{5,n} \cup \mathscr{F}^c_{6,n} \cup \mathscr{F}^c_{7,n}\right\} \to 0$ I obtain

$$\left| \sup_{\alpha_0 \in \mathscr{A}^{(2)}_{\ell_0}(s_0)} \mathbb{P}\left\{ \frac{\sqrt{n}g'(\hat{a}(\widehat{\tau}) - \alpha_0)}{\sqrt{g'\widehat{\Theta}(\widehat{\tau})\hat{\Sigma}(\widehat{\tau})_{xu}\widehat{\Theta}(\widehat{\tau})'g}} \leq t \right\} - \Phi(t) \right| \to 0 \tag{A.6.25}$$

Thus (A.6.7) yields

$$\left| \sup_{\alpha_0 \in \mathscr{B}_{\ell_0}(s_0)} \mathbb{P}\left\{ \frac{\sqrt{n}g'(\hat{a}(\widehat{\tau}) - \alpha_0)}{\sqrt{g'\widehat{\Theta}(\widehat{\tau})\hat{\Sigma}(\widehat{\tau})_{xu}\widehat{\Theta}(\widehat{\tau})'g}} \leq t \right\} - \Phi(t) \right| \to 0 \tag{A.6.26}$$

To see (1.4.31),

$$\begin{aligned}
&\mathbb{P}\left\{ \alpha_0^{(j)} \notin \left[\widehat{a}^{(j)}(\widehat{\tau}) - z_{1-\alpha/2}\frac{\widehat{\sigma}(\widehat{\tau})_j}{\sqrt{n}}, \widehat{a}^{(j)}(\widehat{\tau}) + z_{1-\alpha/2}\frac{\widehat{\sigma}(\widehat{\tau})_j}{\sqrt{n}} \right] \right\} \\
=&\mathbb{P}\left\{ \left| \frac{\sqrt{n}(\widehat{a}^{(j)}(\widehat{\tau}) - \alpha_0^{(j)})}{\widehat{\sigma}^{(j)}} \right| > z_{1-\alpha/2} \right\} \\
=&\mathbb{P}\left\{ \frac{\sqrt{n}(\widehat{a}^{(j)}(\widehat{\tau}) - \alpha_0^{(j)})}{\widehat{\sigma}^{(j)}} > z_{1-\alpha/2} \right\} + \mathbb{P}\left\{ \frac{\sqrt{n}(\widehat{a}^{(j)}(\widehat{\tau}) - \alpha_0^{(j)})}{\widehat{\sigma}^{(j)}} < -z_{1-\alpha/2} \right\} \\
\leq& 1 - \mathbb{P}\left\{ \frac{\sqrt{n}(\widehat{a}^{(j)}(\widehat{\tau}) - \alpha_0^{(j)})}{\widehat{\sigma}^{(j)}} \leq z_{1-\alpha/2} \right\} + \mathbb{P}\left\{ \frac{\sqrt{n}(\widehat{a}^{(j)}(\widehat{\tau}) - \alpha_0^{(j)})}{\widehat{\sigma}^{(j)}} < -z_{1-\alpha/2} \right\}
\end{aligned} \tag{A.6.27}$$

Thus, taking the supremum over $\sup_{\alpha_0 \in \mathscr{B}_{\ell_0}(s_0)}$ and letting n tend to infinity yields an inequality in (1.4.31) via (1.4.30).

Finally turn to (1.4.32), by Lemma 17 and 23 I know

$$\sup_{\alpha_0 \in \mathscr{A}^{(1)}_{\ell_0}(s_0)} |g'\widehat{\Theta}(\hat{\tau})\widehat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g - g'\Theta(\widehat{\tau})\Sigma(\widehat{\tau})_{xu}\Theta(\widehat{\tau})'g| = o_p(1),$$

and

$$\sup_{\alpha_0 \in \mathscr{A}^{(2)}_{\ell_0}(s_0)} |g'\widehat{\Theta}(\hat{\tau})\widehat{\Sigma}(\hat{\tau})_{xu}\widehat{\Theta}(\hat{\tau})'g - g'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'g| = o_p(1).$$

Hence, choosing $g = e_j$ and $\phi_{\max}(\Theta(\tau)) = 1/\phi_{\min}(\Sigma(\tau))$ for $\tau \in \mathbb{T}$,

$$\begin{aligned}
&\sup_{\alpha_0 \in \mathscr{A}^{(1)}_{\ell_0}(s_0)} \operatorname{diam}\left[\widehat{a}^{(j)}(\widehat{\tau}) - z_{1-\alpha/2}\frac{\widehat{\sigma}(\widehat{\tau})_j}{\sqrt{n}}, \widehat{a}^{(j)}(\widehat{\tau}) + z_{1-\alpha/2}\frac{\widehat{\sigma}(\widehat{\tau})_j}{\sqrt{n}}\right] = \sup_{\alpha_0 \in \mathscr{A}^{(2)}_{\ell_0}(s_0)} 2\widehat{\sigma}^{(j)} z_{1-\alpha/2}/\sqrt{n} \\
\text{(A.6.28)} \quad &= 2\left(\sup_{\alpha_0 \in \mathscr{A}^{(1)}_{\ell_0}(s_0)} \sqrt{e_j'\Theta(\widehat{\tau})\Sigma(\widehat{\tau})_{xu}\Theta(\widehat{\tau})'e_j} + o_p(1)\right) z_{1-\alpha/2}/\sqrt{n} \\
&\le 2\left(\sqrt{\phi_{\max}(\Theta(\widehat{\tau}))}\frac{1}{\phi_{\min}(\Sigma(\widehat{\tau}))} + o_p(1)\right) z_{1-\alpha/2}/\sqrt{n} = O_p(1/\sqrt{n})
\end{aligned}$$

$$\begin{aligned}
&\sup_{\alpha_0 \in \mathscr{A}^{(2)}_{\ell_0}(s_0)} \operatorname{diam}\left[\widehat{a}^{(j)}(\widehat{\tau}) - z_{1-\alpha/2}\frac{\widehat{\sigma}(\widehat{\tau})_j}{\sqrt{n}}, \widehat{a}^{(j)}(\widehat{\tau}) + z_{1-\alpha/2}\frac{\widehat{\sigma}(\widehat{\tau})_j}{\sqrt{n}}\right] = \sup_{\alpha_0 \in \mathscr{A}^{(2)}_{\ell_0}(s_0)} 2\widehat{\sigma}^{(j)} z_{1-\alpha/2}/\sqrt{n} \\
\text{(A.6.29)} \quad &= 2\left(\sup_{\alpha_0 \in \mathscr{A}^{(2)}_{\ell_0}(s_0)} \sqrt{e_j'\Theta(\tau_0)\Sigma(\tau_0)_{xu}\Theta(\tau_0)'e_j} + o_p(1)\right) z_{1-\alpha/2}/\sqrt{n} \\
&\le 2\left(\sqrt{\phi_{\max}(\Theta(\tau_0))}\frac{1}{\phi_{\min}(\Sigma(\tau_0))} + o_p(1)\right) z_{1-\alpha/2}/\sqrt{n} = O_p(1/\sqrt{n})
\end{aligned}$$

Therefore, (1.4.32) is proven. □

A.7 Time Series Model

While I develop the theory in the context of independent data, I also explain how the theory remains applicable in time series data models with certain assumptions modified.

Assumption 7. *(i)$\{X_i, U_i, Q_i\}_{i=1}^n$ are sequences of (strictly) stationary and ergodic random variables. Furthermore, marginal distribution of $\{Q_i\}_{i=1}^n$ is uniform $(0,1)$ and $\{U_i\}_{i=1}^n$, and $\{X_i\}_{i=1}^n$*

are independent. (ii)For the strong mixing variables X_i, U_i: $\alpha(i) \geq exp(-Ci^{r_0})$, for a positive constant $r_0 > 0$. (iii)There exists positive constants r_1, r_2,and another set of positive constants b_1,b_2,s_1, $s_2 > 0$, and for $i = 1,\cdots,n$, and $j = 1,\cdots,p$ $\mathbb{P}\{|U_i| > s_1\} \leq exp(-(s_1/b_1)^{r_1})$ and $\mathbb{P}\left\{|X_i^{(j)}| > s_2\right\} \leq exp(-(s_2/b_2)^{r_2})$ (iv)There exists $0 < \gamma_1 < 1$ such that $\gamma_1^{-1} = 3r_1^{-1} + r_0^{-1}$ and $3r_2^{-1} + r_0^{-1} > 1$. (v)There exist positive constants r_3 and another set of positive constants $b_3, s_3 > 0$. For $i = 1,\cdots,n$, and $j = 1,\cdots,p$, $\mathbb{P}\left\{|\upsilon_i^{(j)}| > s_3\right\} \leq \exp(-(s_3/b_3)^{r_3})$, and the same γ_1 as in (iv) such that $3r_3^{-1} + r_0^{-1} > 1$.

Assumption 7 is regarded as a modification of Assumption 1, while keeping other assumptions unchanged. It is noteworthy that stationary GARCH models with finite second moments and continuous error distributions, as well as causal ARMA processes with continuous error distributions, and a specific class of stationary Markov chains satisfy my Assumptions 7. Similar assumptions are discussed in Chang et al. (2018) and Caner et al. (2023).

The following is Lemma A.3(i) of Fan et al. (2011) under Assumption 7:

$$\mathbb{P}\left\{\max_{1\leq j\leq p}\max_{1\leq l\leq p}|\frac{1}{n}\sum_{i=1}^{n}X_i^{(j)}X_i^{(l)} - E[(X_i^{(j)}X_i^{(l)}]| \geq C\frac{\sqrt{\log p}}{\sqrt{n}}\right\} = O(\frac{1}{p^2}),$$

The following is Lemma B.1(ii) of Fan et al. (2011) under Assumption 7:

$$\mathbb{P}\left\{\max_{1\leq j\leq p}|\frac{1}{n}\sum_{i=1}^{n}U_iX_i^{(j)}| \geq C\frac{\sqrt{\log p}}{\sqrt{n}}\right\} = O(\frac{1}{p^2})$$

The proof of Lemma A.3(i) and Lemma B.1(ii) in Fan et al. (2011) relies on the maximal inequality presented in Lemma A.2 of the same reference, attributed to Theorem 1 in Merlevède et al. (2011).

Apply the same techniques employed in the proofs uniformly over $\tau \in \mathbb{T}$, incorporating an additional layer as specified by (A.1.15), I can show

$$\mathbb{P}\left\{\max_{1\leq j\leq p}\sup_{\tau\in\mathbb{T}}\frac{1}{n}\sum_{i=1}^{n}U_iX_i^{(j)}\mathbf{1}\{Q_i<\tau\}\geq C\frac{\sqrt{\log np}}{\sqrt{n}}\right\}=O(\frac{1}{n^2p^2})$$

$$\mathbb{P}\left\{\sup_{\tau\in\mathbb{T}}\max_{1\leq j\leq p}\max_{1\leq l\leq p}|\frac{1}{n}\sum_{i=1}^{n}X_i^{(j)}X_i^{(l)}\mathbf{1}\{Q_i<\tau\}-E[X_i^{(j)}X_i^{(l)}\mathbf{1}\{Q_i<\tau\}]|\geq C\frac{\sqrt{\log np}}{\sqrt{n}}\right\}=O(\frac{1}{n^2p^2}),$$

By substituting the maximal inequality from Lemma A.2 in Fan et al. (2011), under Assumption 7, in place of the inequalities from Lemma E.1 and E.2 of Chernozhukov et al. (2017) used in all previous proofs, I can establish that $\mathbb{P}\{\mathbb{A}_1\}$, $\mathbb{P}\{\mathbb{A}_2\}$, $\mathbb{P}\{\mathbb{A}_3\}$, $\mathbb{P}\{\mathbb{A}_4\}$, and $\mathbb{P}\{\mathbb{A}_5\}$ approach 1 for all sufficiently large n and $p>n$. These results imply that my framework encompasses the time series data threshold regression model.

A.8 Threshold selection consistency by thresholding

In the case of a linear model, van de Geer et al. (2014) has already addressed debiased LASSO estimation for uniformly valid confidence bands. However, when dealing with a well-identified and discontinuous threshold effect, I need to propose a debiased LASSO. It is crucial to determine whether a threshold is present or absent, even in the context of high-dimensional threshold models with random regressors. But econometricians do not have prior knowledge of whether a threshold is present. Precise variable selection becomes crucial. As pointed out by Callot et al. (2017), a sup-norm bound provides more accurate variable selection results for the thresholded scaled LASSO compared to results based on ℓ_1 bounds. The latter tends to be larger due to the presence of the unknown sparsity s_0.Up to this point, I have established oracle inequalities for the prediction norm and ℓ_1 errors of my estimates. Before delving into the desparsification of the estimator for test and confidence interval construction, I address the threshold detection issue.

The situation where $\delta_0=0$ is non-trivial since the consistency of an estimator does not provide selection consistency. Suppose $\delta_0 = 0$, Theorem 1 shows that

$$\widehat{\delta}^{(j)}(\widehat{\tau}) \xrightarrow{p} 0,$$

for each $j \in \{1, \cdots p\}$. However, this does not imply that I will correctly estimate zero coefficients as zero. The consistency implies that for all $\varepsilon > 0$,

$$\mathbb{P}\left\{|\widehat{\delta}^{(j)}(\widehat{\tau})| \geq \varepsilon\right\} \to 0$$

But as I need to control the correct model, I instead require

$$\mathbb{P}\left\{\widehat{\delta}^{(j)}(\widehat{\tau}) = 0\right\} \to 1. \tag{A.8.1}$$

(A.8.1) states that, with a consistent estimator, selection consistency comes from (A.8.1). In particular, LASSO has a tendency to overshoot the correct model, finding more nonzero coefficients than the true number. Strictly speaking, if the estimated number of nonzero coefficients is $\widehat{s}$, then in finite samples LASSO has a tendency to obtain $\widehat{s} > s_0$. To my scaled threshold model, the LASSO estimator defined in (1.2.4) may be much more over-parameterized in that τ and δ are added to β as parameters.

We next turn to variable selection by means of thresholding. For this purpose, I follow Callot et al. (2017) to define the thresholded LASSO estimator[1] as

$$\tilde{\delta}^{(j)}(\widehat{\tau}) = \begin{cases} \widehat{\delta}^{(j)}(\widehat{\tau}), & \text{if } |\widehat{\delta}^{(j)}(\widehat{\tau})| \geq H, \\ 0, & \text{if } |\widehat{\delta}^{(j)}(\widehat{\tau})| < H. \end{cases} \tag{A.8.2}$$

where H is the threshold determining whether a coefficient should be classified as zero or nonzero and $\widehat{\delta}^{(j)}(\widehat{\tau})$ are elements of the LASSO estimator defined by (1.2.4). In particular, I shall see that choosing $H = 2C\lambda$ results in consistent model selection.

Theorem 5 (Threshold selection consistency)**.** *Let Assumptions 1-4 hold and assume that*

[1] Note that since I are only interested in finding out whether δ_0 is nonzero or not, one can simply threshold $\widehat{\delta}$.

$\min_{j\in J(\delta_0)}|\delta_0^{(j)}| > 3C\lambda$. *Then* $\forall\varepsilon > 0$, *there exists a* C *such that for* $H = 2C\lambda = 2C\sqrt{\frac{\log p}{n}}$,

$$\mathbb{P}\left\{J(\delta_0) = J(\tilde{\delta}(\hat{\tau}))\right\} \geq 1-\varepsilon$$

as $n \to \infty$.

Theorem 5 is derived from Theorem 4 in Callot et al. (2017). The discussion on choosing the thresholding parameter C through the Bayesian Information Criterion (BIC) is omitted, as it is similarly implemented in the simulation section of Callot et al. (2017). Theorem 5 outlines sufficient conditions for the thresholded LASSO to identify the correct sparsity pattern of δ_0. It is essential to highlight that these conditions require the absolute value of the smallest non-zero coefficient to be at least of the order of the ℓ_∞-rate of convergence of $\hat{\alpha}$ to α_0.

There exists a trade-off in deciding whether to include this assumption. If included, the construction of confidence bands for parameters doesn't yield uniformly valid results over any ℓ_0-ball $\mathscr{B}(s_0)$, as the result relies on $\min_{j\in J(\delta_0)}|\delta_0^{(j)}| > 3C\lambda$ for validity extends beyond the complement of every such ℓ_0-ball.

On the other hand, without this assumption, the condition $\tilde{\delta}(\hat{\tau}) \neq 0$ is sufficient to imply the true model is nonlinear. However, the condition $\tilde{\delta}(\hat{\tau}) = 0$ is not sufficient to imply the true model is linear.

A.9 Asymptotic Distribution of Threshold Parameter

To develop the asymptotic properties of the threshold parameter estimator, I rely on the empirical process results introduced by Hansen (2000) and adopt the shrinking-threshold-effect framework. In this framework, the threshold effect diminishes as the sample size tends to infinity. By constructing a likelihood ratio (LR) statistic, I can derive inferences regarding the threshold parameter.

Assumption 8. *(i) For some fixed* $\delta_0^* < \infty$ *and* $0 < \varphi < \frac{1}{2}$, *let* $\delta_0 = n^{-\varphi}\delta_0^*$ *and* $n^{-\varphi}\|\delta_0^*\|_1 > 0$.
(ii) $E\left[X_i X_i' U_i^2 | Q_i = \tau\right]$ *is continuous and bounded when* τ *is in a neighborhood of* τ_0.

(iii) For any $\eta > 0$ and $\tau_1, \tau_2 \in \mathbb{T}$ such that wpa1,

$$\text{(A.9.1)} \qquad \sup_{|\tau-\tau_0|<\eta} \left| \frac{1}{n} \sum_{i=1}^{n} U_i X_i' \delta_0 [1(Q_i < \tau_0) - 1(Q_i < \tau)] \right| \leq \frac{\lambda(\sqrt{\eta})^{\varpi}}{2},$$

$$\text{(A.9.2)} \qquad \sup_{1 \leq j,l \leq p} \sup_{|\tau-\tau_0|<\eta} \frac{1}{n} \sum_{i=1}^{n} \left| X_i^{(j)} X_i^{(l)} \right| |1(Q_i < \tau_0) - 1(Q_i < \tau)| \leq C_5(\eta)^{\varpi},$$

$$\text{(A.9.3)} \qquad \sup_{1 \leq j,l \leq p} \sup_{|\tau-\tau_0|<\eta} \|\delta_0\|_1 \left| \frac{1}{n} \sum_{i=1}^{n} U_i X_i^{(j)} [1(Q_i < \tau_0) - 1(Q_i < \tau)] \right| \leq \frac{\lambda(\sqrt{\eta})^{\varpi}}{2},$$

where $0 < \varphi < \frac{1}{2}$, $\varpi > \frac{1}{1-2\varphi}$.

(iv) s_0, ϖ, φ, n, and p are such that $\frac{s_0^2 \log p}{n^{\varpi(1-2\varphi)-1}} \|\delta_0\|_1^2 = o_p(1)$, $\sqrt{\frac{\log p}{n^{1-2\varphi}}} = o_p(1)$, $\|\delta_0\|_1^4 \frac{\log p}{n} = o_p(1)$.

Assumption 8 is an extension of the fixed dimension case in the literature when working with a fixed regressor design(e.g., Hansen (2000)). Assumption 8 (i) has been widely used in the threshold model to obtain a tractable asymptotic distribution for the least squares estimator of τ (e.g., Hansen (2000)) The re-normalization is to force δ_0 to be small, reducing the information in the sample concerning the threshold and hence slowing down the rate of convergence of the threshold estimate. This assumption need not be viewed as very restrictive since the rate at which δ_0 decreases to zero can be set quite low. It does suggest, however, that the asymptotic approximation is more likely to provide good approximations when δ_0 is small relative to the case where δ_0 is large. The unknown parameter $0 < \varphi < \frac{1}{2}$ reflects the difficulty of estimating and affects the identification and estimation of the change point. Both the rate of convergence and the asymptotic distribution depend on. In Assumption 8 (iii), (A.9.3) implies (A.9.1).

The following arguments are parallel to those in Lemma 11, Lemma 13, Theorem 1, and Theorem 2 of Hansen (2000). To describe the asymptotic distribution, I introduce additional notations. For any $v \in \Psi$, an arbitrary compact set, let

$$\Delta_n(v) = n\left[S_n(\widehat{\alpha}(\widehat{\tau}), \tau_0) + \lambda \|\boldsymbol{D}(\tau_0)\widehat{\alpha}(\widehat{\tau})\|_1\right] - n\left[S_n(\widehat{\alpha}(\widehat{\tau}), \tau_0 + \frac{v}{n^{1-2\varphi}}) + \lambda \left\|\boldsymbol{D}(\tau_0 + \frac{v}{n^{1-2\varphi}})\widehat{\alpha}(\widehat{\tau})\right\|_1\right]$$

Let $\widehat{v} = n^{1-2\varphi}(\widehat{\tau} - \tau_0)$, I can then derive the process using (1.2.4)

(A.9.4)

$$
\begin{aligned}
&\operatorname{argmax}_v \Delta_n(v)\\
=&\operatorname{argmax}_v S_n(\widehat{\alpha}(\widehat{\tau}), \tau_0) + \lambda \|\boldsymbol{D}(\tau_0)\widehat{\alpha}(\widehat{\tau})\|_1 - S_n(\widehat{\alpha}(\widehat{\tau}), \tau_0 + \frac{v}{n^{1-2\varphi}}) - \lambda \left\|\boldsymbol{D}(\tau_0 + \frac{v}{n^{1-2\varphi}})\widehat{\alpha}(\widehat{\tau})\right\|_1\\
=&\widehat{v},
\end{aligned}
$$

as $S_n(\widehat{\alpha}(\widehat{\tau}), \widehat{\tau}) + \lambda \|\boldsymbol{D}(\widehat{\tau})\widehat{\alpha}(\widehat{\tau})\|_1$ represents the minimum over $\tau \in \mathbb{T}$.

$$
\begin{aligned}
&\Delta_n(v)\\
=&\widehat{\delta}(\widehat{\tau})'(X(\tau_0 + \frac{v}{n^{1-2\varphi}}) - X(\tau_0))'(X(\tau_0 + \frac{v}{n^{1-2\varphi}})\widehat{\delta}(\widehat{\tau}) - 2\widehat{\delta}(\widehat{\tau})(X(\tau_0 + \frac{v}{n^{1-2\varphi}}) - X(\tau_0))U\\
&+2\widehat{\delta}(\widehat{\tau})(X(\tau_0 + \frac{v}{n^{1-2\varphi}}) - X(\tau_0))'(X(\tau_0 + \frac{v}{n^{1-2\varphi}})(\widehat{\beta}(\widehat{\tau}) - \beta_0))\\
&+n\lambda \left\|\boldsymbol{D}(\tau_0 + \frac{v}{n^{1-2\varphi}})\alpha(\widehat{\tau})\right\|_1 - n\lambda \|\boldsymbol{D}(\tau_0)\alpha(\widehat{\tau})\|_1\\
=&\delta_0'(X(\tau_0 + \frac{v}{n^{1-2\varphi}}) - X(\tau_0))'X(\tau_0 + \frac{v}{n^{1-2\varphi}})\delta_0 - 2\delta_0'(X(\tau_0 + \frac{v}{n^{1-2\varphi}}) - X(\tau_0))'U\\
&-2(\widehat{\delta}(\widehat{\tau}) - \delta_0)'(X(\tau_0 + \frac{v}{n^{1-2\varphi}}) - X(\tau_0))'U\\
&+2\widehat{\delta}(\widehat{\tau})'(X(\tau_0 + \frac{v}{n^{1-2\varphi}}) - X(\tau_0))'X(\tau_0 + \frac{v}{n^{1-2\varphi}})(\widehat{\beta}(\widehat{\tau}) - \beta_0)\\
&+(\widehat{\delta}(\widehat{\tau}) + \delta_0)'(X(\tau_0 + \frac{v}{n^{1-2\varphi}}) - X(\tau_0))'X(\tau_0 + \frac{v}{n^{1-2\varphi}})(\widehat{\delta}(\widehat{\tau}) - \delta_0)\\
&+n\lambda \left\|\boldsymbol{D}(\tau_0 + \frac{v}{n^{1-2\varphi}})\alpha(\widehat{\tau})\right\|_1 - n\lambda \|\boldsymbol{D}(\tau_0)\alpha(\widehat{\tau})\|_1
\end{aligned}
\tag{A.9.5}
$$

Regarding the second term in the last equation in (A.9.5), I introduce additional notations. Let $R_n(v) = \frac{\sqrt{n^{1-2\varphi}}}{\sqrt{n}} \sum_{i=1}^n \delta_0^{*\prime}(X_i(\tau_0 + \frac{v}{n^{1-2\varphi}}) - X_i(\tau_0))U_i$ and $V_n(v) = \frac{n^{1-2\varphi}}{n} \sum_{i=1}^n \delta_0^{*\prime}(X_i(\tau_0 + \frac{v}{n^{1-2\varphi}}) - X_i(\tau_0))(X_i(\tau_0 + \frac{v}{n^{1-2\varphi}}) - X_i(\tau_0))'\delta_0^* U_i^2$. First, I show for any given v the convergence of the finite-dimensional distributions of $R_n(v)$ to those of $B(v)$. It suffices to show the first and second terms in the last equation in (A.9.5) converge somewhere correspondingly, and then show the convergence of $\Delta_n(v)$.

Lemma 24. *Under Assumption 1,2 and 8, for any $v \in \Psi$, a arbitrary compact set,*

$$
\delta_0'(X(\tau_0 + \frac{v}{n^{1-2\varphi}}) - X(\tau_0))'X(\tau_0 + \frac{v}{n^{1-2\varphi}})\delta_0 \xrightarrow{p} v\delta_0^{*\prime} E\left[X_i X_i' | Q_i = \tau_0\right]\delta_0^*
$$

and

$$R_n(v) \rightsquigarrow B(v),$$

where $B(v)$ *can be written as* $\sqrt{\delta_0^{*\prime} E\left[X_i X_i' U_i^2 | Q_i = \tau\right] \delta_0^*}\, W(v)$, *and* $W(v)$ *is a standard Brownian motion.*

The notation $R_n(v) \rightsquigarrow B(v)$ defines a general concept of convergence in distribution introduced by Dudley (1985).

Proof. To show the first part of the lemma,

$$\begin{aligned}
& E\left[\delta_0'(X(\tau_0+\frac{v}{n^{1-2\varphi}})-X(\tau_0))'X(\tau_0+\frac{v}{n^{1-2\varphi}})\delta_0\right] \\
=&\delta_0' E\left[(X(\tau_0+\frac{v}{n^{1-2\varphi}})'X(\tau_0+\frac{v}{n^{1-2\varphi}})-X(\tau_0)'(X(\tau_0+\frac{v}{n^{1-2\varphi}})\right]\delta_0 \\
=&\frac{n^{1-2\varphi}}{n}\delta_0^{*\prime} E\left[X(\tau_0+\frac{v}{n^{1-2\varphi}})'X(\tau_0+\frac{v}{n^{1-2\varphi}})-X(\tau_0)'X(\tau_0)\right]\delta_0^* \\
=&n^{1-2\varphi}\delta_0^{*\prime} E\left[\frac{1}{n}X(\tau_0+\frac{v}{n^{1-2\varphi}})'X(\tau_0+\frac{v}{n^{1-2\varphi}})-X(\tau_0)'X(\tau_0)\right]\delta_0^* \\
=&n^{1-2\varphi}\delta_0^{*\prime} E\left[\frac{1}{n}\sum_{i=1}^{n} X_i X_i'\left[\mathbf{1}\left(Q_i<\tau_0+\frac{v}{n^{1-2\varphi}}\right)-\mathbf{1}(Q_i<\tau_0)\right]U_i^2\right]\delta_0^* \\
=&n^{1-2\varphi}\delta_0^{*\prime} E\left[X_i X_i' | \tau_0 \le Q_i \le \tau_0+\frac{v}{n^{1-2\varphi}}\right]\delta_0^* \\
\xrightarrow{p}& v\delta_0^{*\prime} E\left[X_i X_i' | Q_i=\tau_0\right]\delta_0^*
\end{aligned} \tag{A.9.6}$$

as $n \to \infty$.

$$\begin{aligned}
& E\left[\delta_0'(X(\tau_0+\frac{v}{n^{1-2\varphi}})-X(\tau_0))'(X(\tau_0+\frac{v}{n^{1-2\varphi}})\delta_0-E\left[\delta_0'(X(\tau_0+\frac{v}{n^{1-2\varphi}})-X(\tau_0))'X(\tau_0+\frac{v}{n^{1-2\varphi}})\delta_0\right]\right]^2 \\
=&E\left[\delta_0'\left[X(\tau_0+\frac{v}{n^{1-2\varphi}})'X(\tau_0+\frac{v}{n^{1-2\varphi}})-E\left[X(\tau_0+\frac{v}{n^{1-2\varphi}})'X(\tau_0+\frac{v}{n^{1-2\varphi}})\right]-X(\tau_0)'X(\tau_0)+E\left[X(\tau_0)'X(\tau_0)\right]\right]\delta_0\right]^2 \\
\le&E\left[\|\delta_0\|_1^2\left[\|\frac{1}{n}X_i(\tau_0+v)'X_i(\tau_0+v)-E\left[X_i(\tau_0+v)'X_i(\tau_0+v)\right]\|_\infty+\|\frac{1}{n}X_i(\tau_0)'X_i(\tau_0)-E\left[X_i(\tau_0)'X_i(\tau_0)\right]\|_\infty\right]\right]^2 \\
\xrightarrow{p}&\|\delta_0\|_1^4 O_p(\frac{\log p}{n})
\end{aligned}$$

where I used Lemma 6 in the last step. Combine the above with Markov's inequality,

$$\delta_0'(X(\tau_0+\frac{v}{n^{1-2\varphi}})-X(\tau_0))'X(\tau_0+\frac{v}{n^{1-2\varphi}})\delta_0 \xrightarrow{p} v\delta_0^{*\prime} E\left[X_i X_i' | Q_i=\tau_0\right]\delta_0^*.$$

My proof proceeds by establishing the convergence of the finite-dimensional distributions of $R_n(v)$ to those of $B(v)$ for any given v, then extending that by showing the tightness of $R_n(v)$.

$$
\begin{aligned}
&E\left[V_n(v)\right] \\
=&E\left[\frac{n^{1-2\varphi}}{n}\sum_{i=1}^{n}\delta_0^{*\prime}X_iX_i'\delta_0^*\left[\mathbf{1}\left(Q_i<\tau_0+\frac{v}{n^{1-2\varphi}}\right)-\mathbf{1}\left(Q_i<\tau_0\right)\right]U_i^2\right] \\
=&\sum_{i=1}^{n}E\left[\frac{n^{1-2\varphi}}{n}\delta_0^{*\prime}X_iX_i'\delta_0^*\left[\mathbf{1}\left(Q_i<\tau_0+\frac{v}{n^{1-2\varphi}}\right)-\mathbf{1}\left(Q_i<\tau_0\right)\right]U_i^2\right] \\
\xrightarrow{p}&v\delta_0^{*\prime}E\left[X_iX_i'U_i^2|Q_i=\tau\right]\delta_0^*
\end{aligned}
\tag{A.9.7}
$$

$$
\begin{aligned}
&E\left[V_n(v)-E[V_n(v)]\right]^2 \\
=&E\left[\frac{n^{1-2\varphi}}{n}\sum_{i=1}^{n}\delta_0^{*\prime}X_iX_i'\delta_0^*[1(Q_i<\tau_0)-1(Q_i<\tau)]U_i^2-E\left[\frac{n^{1-2\varphi}}{n}\sum_{i=1}^{n}\delta_0^{*\prime}X_iX_i'\delta_0^*[1(Q_i<\tau_0)-1(Q_i<\tau)]U_i^2\right]\right] \\
=&n^{1-2\varphi}E\left[\frac{1}{n}\sum_{i=1}^{n}\delta_0^{*\prime}X_iX_i'\delta_0^*[1(Q_i<\tau_0)-1(Q_i<\tau)]U_i^2-E\left[\frac{1}{n}\sum_{i=1}^{n}\delta_0^{*\prime}X_iX_i'\delta_0^*[1(Q_i<\tau_0)-1(Q_i<\tau)]U_i^2\right]\right] \\
=&n^{1-2\varphi}E\left[\frac{1}{n}\sum_{i=1}^{n}\delta_0^{*\prime}X_iX_i'\delta_0^*[1(Q_i<\tau_0)-1(Q_i<\tau)]U_i^2-E\left[\frac{1}{n}\sum_{i=1}^{n}\delta_0^{*\prime}X_iX_i'\delta_0^*[1(Q_i<\tau_0)-1(Q_i<\tau)]U_i^2\right]\right] \\
\leq&n^{1-2\varphi}E\left[\|\delta_0^*\|_1^2\left[\|\frac{1}{n}X_i(\tau_0+v)'X_i(\tau_0+v)-E\left[X_i(\tau_0+v)'X_i(\tau_0+v)\right]\|_\infty+\|\frac{1}{n}X_i(\tau_0)'X_i(\tau_0)-E\left[X_i(\tau_0)'X_i(\tau_0)\right]\|_\infty\right]\right] \\
\xrightarrow{p}&0
\end{aligned}
$$

which establishes that $V_n(v)\xrightarrow{p}|v|\delta_0^{*\prime}E\left[X_iX_i'U_i^2|Q_i=\tau\right]\delta_0^*$ by Markov's inequality.

Since $\{X_i,U_i,Q_i\}_{i=1}^n$ is an independent and identically distributed sequence, $E[R_n(v)]=0$. We conclude that $R_n(v)\xrightarrow{d}N(0,|v|\delta_0^{*\prime}E\left[X_iX_i'U_i^2|Q_i=\tau\right]\delta_0^*)$ for any fix v. This argument can be extended to include any finite collection $[v_1,\cdots v_k]$, to yield the convergence of the finite-dimensional distributions of $R_n(v)$ to those of $B(v)$.

Then I show the tightness of $R_n(v)$. Fix $\eta>0$ and set $\tau_1=\tau_0+\frac{v_1}{n^{1-2\varphi}}$, then by Assumption 8(A.9.1),

$$
\sup_{v_1\leq v\leq v_1+\eta}R_n(v)-R_n(v_1)\leq\sup_{v_1\leq v\leq v_1+\eta}R_n(v)-R_n(0)+\sup_{v_1\leq v\leq v_1+\eta}R_n(0)-R_n(v_1)\leq\frac{1}{2}\lambda n^{1-\varpi(\frac{1}{2}-\varphi)}[(\sqrt{v})^{\varpi}+(\sqrt{v_1})^{\varpi}].
$$

Thus, $\mathbb{P}\{\sup_{v_1\leq v\leq v_1+\eta}|R_n(v)-R_n(v_1)|>\lambda n^{1-\frac{1}{2}\varpi(1-2\varphi)}(\sqrt{v})^{\varpi}\}\to 0$, as $n\to\infty$. So $R_n(v)$ is tight. As $R_n(v)$ is tight, I conclude that $R_n(v)\rightsquigarrow B(v)$. $\square$

In next part, I present an auxiliary technical lemma and its proof. We start with some matrix norm inequalities. Let A be a generic $q\times p$ matrix and x a $p\times 1$ vector and z a $q\times 1$ vector.

Lemma 25.

$$x'Az \leq \|x\|_1 \|A\|_\infty \|z\|_1$$

Proof. Observe that

$$x'Az \leq \|x\|_1 \|Az\|_\infty \leq \|x\|_1 \|A\|_\infty \|z\|_1$$

by Hölder's inequality. □

Lemma 26. *Under Assumption 1,2 and 8, for any $v \in \Psi$,on any compact set,*

$$\Delta_n(v) \rightsquigarrow \Delta(v),$$

where $\Delta(v) = -|v| \delta_0^{*\prime} E\left[X_i X_i' | Q_i = \tau_0\right] \delta_0^* + 2\sqrt{\delta_0^{*\prime} E\left[X_i X_i' U_i^2 | Q_i = \tau\right] \delta_0^*} W(v)$, *and $W(v)$ is a standard Brownian motion.*

Proof. Rearranging (A.9.5), yields

$$\Delta_n(v) = \delta_0'(X(\tau_0 + \frac{v}{n^{1-2\varphi}}) - X(\tau_0)) X(\tau_0 + \frac{v}{n^{1-2\varphi}})' \delta_0 + 2R_n(v) + \Upsilon(v), \tag{A.9.8}$$

where

$$\begin{aligned} \Upsilon(v) = &-2(\widehat{\delta}(\widehat{\tau}) - \delta_0)'(X(\tau_0 + \frac{v}{n^{1-2\varphi}}) - X(\tau_0))U \\ &+2\widehat{\delta}(\widehat{\tau})'(X(\tau_0 + \frac{v}{n^{1-2\varphi}}) - X(\tau_0)) X(\tau_0 + \frac{v}{n^{1-2\varphi}})'(\widehat{\beta}(\widehat{\tau}) - \beta_0) \\ &+(\widehat{\delta}(\widehat{\tau}) + \delta_0)'(X(\tau_0 + \frac{v}{n^{1-2\varphi}}) - X(\tau_0)) X(\tau_0 + \frac{v}{n^{1-2\varphi}})'(\widehat{\delta}(\widehat{\tau}) - \delta_0) \\ &+n\lambda \left\| \boldsymbol{D}(\tau_0 + \frac{v}{n^{1-2\varphi}})\alpha(\widehat{\tau}) \right\|_1 - n\lambda \|\boldsymbol{D}(\tau_0)\alpha(\widehat{\tau})\|_1 \end{aligned} \tag{A.9.9}$$

It suffices to show $\Upsilon(v) \Rightarrow 0$. Note that by triangle inequality and Hölder's inequality

$$
\begin{aligned}
& n\lambda \left\| \boldsymbol{D}(\tau_0 + \frac{v}{n^{1-2\varphi}})\alpha(\widehat{\tau}) \right\|_1 - n\lambda \|\boldsymbol{D}(\tau_0)\alpha(\widehat{\tau})\|_1 \\
\leq & n\lambda \left\| \left(\boldsymbol{D}(\tau_0 + \frac{v}{n^{1-2\varphi}}) - \boldsymbol{D}(\tau_0) \right) \alpha(\widehat{\tau}) \right\|_1 \\
= & n\lambda \left| \sum_{j=1}^{p} \left(\|X^{(j)}(\tau_0 + \frac{v}{n^{1-2\varphi}})\|_n - \|X^{(j)}(\tau_0)\|_n \right) \widehat{\delta}^{(j)}(\widehat{\tau}) \right| \\
\leq & n\lambda \left| \sum_{j=1}^{p} \|X^{(j)}(\tau_0 + \frac{v}{n^{1-2\varphi}}) - X^{(j)}(\tau_0)\|_n \widehat{\delta}^{(j)}(\widehat{\tau}) \right| \\
\leq & n\lambda \left(\max_{j=1\cdots p} \left\| X^{(j)}(\tau_0 + \frac{v}{n^{1-2\varphi}}) - X^{(j)}(\tau_0) \right\|_n \right) \left\| \widehat{\delta}(\widehat{\tau}) \right\|_1 \\
\leq & n\lambda \sqrt{(\frac{v}{n^{1-2\varphi}})^{\varpi}} \left\| \widehat{\delta}(\widehat{\tau}) \right\|_1 \\
\leq & n\lambda \sqrt{(\frac{v}{n^{1-2\varphi}})^{\varpi}} \left(\|\delta_0\|_1 + C s_0 \lambda \right) \\
\leq & n\lambda \sqrt{(\frac{v}{n^{1-2\varphi}})^{\varpi}} \left(n^{-\varphi} \left\| \delta_0^* \right\|_1 + C s_0 \lambda \right) \\
\leq & C v^{\frac{\varpi}{2}} \left(\frac{\sqrt{\log p} \left\| \delta_0^* \right\|_1}{\sqrt{n^{(1-2\varphi)(\varpi-1)}}} + \frac{s_0 \log p}{\sqrt{n^{(1-2\varphi)\varpi}}} \right) \\
= & O_p \left(\frac{\sqrt{\log p} \left\| \delta_0^* \right\|_1}{\sqrt{n^{(1-2\varphi)(\varpi-1)}}} + \frac{s_0 \log p}{\sqrt{n^{(1-2\varphi)\varpi}}} \right)
\end{aligned} \tag{A.9.10}
$$

where the last inequality is by Assumption 8 (A.9.2), Theorem 1 or 2.

Note that by Hölder's inequality and Assumption 8 (A.9.3)

$$
\begin{aligned}
& 2(\widehat{\delta}(\widehat{\tau}) - \delta_0)'(X(\tau_0 + \frac{v}{n^{1-2\varphi}}) - X(\tau_0))U \\
\leq & 2\|\widehat{\delta}(\widehat{\tau}) - \delta_0\|_1 \sup_{1\leq j\leq p} \sup_{|\tau-\tau_0|<\frac{v}{n^{1-2\varphi}}} \left| \sum_{i=1}^{n} U_i X_i^{(j)} [1(Q_i < \tau_0) - 1(Q_i < \tau)] \right| \\
\leq & \|\widehat{\delta}(\widehat{\tau}) - \delta_0\|_1 n \frac{\lambda (\sqrt{\frac{v}{n^{1-2\varphi}}})^{\varpi}}{\|\delta_0\|_1} \\
\leq & C \frac{s_0 \log p}{n^{-\varphi} \|\delta_0^*\|_1} \sqrt{(\frac{v}{n^{1-2\varphi}})^{\varpi}} \\
= & C v^{\frac{\varpi}{2}} \frac{s_0 \log p}{\|\delta_0^*\|_1 \sqrt{n^{(1-2\varphi)\varpi - 2\varpi}}} \\
= & O_p \left(\frac{s_0 \log p}{\|\delta_0^*\|_1 \sqrt{n^{(1-2\varphi)\varpi - 2\varpi}}} \right).
\end{aligned} \tag{A.9.11}
$$

By Lemma 25 and Assumption 8 (A.9.2)

$$\begin{aligned}
&2\widehat{\delta}(\widehat{\tau})'(X(\tau_0+\frac{v}{n^{1-2\varphi}})-X(\tau_0))'X(\tau_0+\frac{v}{n^{1-2\varphi}})(\widehat{\beta}(\widehat{\tau})-\beta_0)\\
=&2(\widehat{\delta}(\widehat{\tau})'-\delta_0'+\delta_0')(X(\tau_0+\frac{v}{n^{1-2\varphi}})-X(\tau_0))'X(\tau_0+\frac{v}{n^{1-2\varphi}})(\widehat{\beta}(\widehat{\tau})-\beta_0)\\
\le&2\|\widehat{\delta}(\widehat{\tau})-\delta_0\|_1\sup_{1\le j,l\le p}\sup_{|\tau-\tau_0|<\frac{v}{n^{1-2\varphi}}}\left|\sum_{i=1}^n X_i^{(j)}X_i^{(l)}[1(Q_i<\tau_0)-1(Q_i<\tau)]\right|\|\widehat{\beta}(\widehat{\tau})-\beta_0\|_1\\
&+2\|\delta_0\|_1\sup_{1\le j,l\le p}\sup_{|\tau-\tau_0|<\frac{v}{n^{1-2\varphi}}}\left|\sum_{i=1}^n X_i^{(j)}X_i^{(l)}[1(Q_i<\tau_0)-1(Q_i<\tau)]\right|\|\widehat{\beta}(\widehat{\tau})-\beta_0\|_1\\
\le&C(\frac{v}{n^{1-2\varphi}})^{\varpi}n(\frac{s_0\sqrt{\log p}}{\sqrt{n}})^2+C(\frac{v}{n^{1-2\varphi}})^{\varpi}n\frac{s_0\sqrt{\log p}}{\sqrt{n}}\|\delta_0\|_1
\end{aligned}\tag{A.9.12}$$

By Lemma 25 and Assumption 8 (A.9.2)

$$\begin{aligned}
&(\widehat{\delta}(\widehat{\tau})+\delta_0)'(X(\tau_0+\frac{v}{n^{1-2\varphi}})-X(\tau_0))X(\tau_0+\frac{v}{n^{1-2\varphi}})'(\widehat{\delta}(\widehat{\tau})-\delta_0)\\
=&(\widehat{\delta}(\widehat{\tau})-\delta_0+2\delta_0)'(X(\tau_0+\frac{v}{n^{1-2\varphi}})-X(\tau_0))X(\tau_0+\frac{v}{n^{1-2\varphi}})'(\widehat{\delta}(\widehat{\tau})-\delta_0)\\
\le&\|\widehat{\delta}(\widehat{\tau})-\delta_0\|_1\sup_{1\le j,l\le p}\sup_{|\tau-\tau_0|<\frac{v}{n^{1-2\varphi}}}\left|\sum_{i=1}^n X_i^{(j)}X_i^{(l)}[1(Q_i<\tau_0)-1(Q_i<\tau)]\right|\|\widehat{\delta}(\widehat{\tau})-\delta_0\|_1\\
&+2\|\delta_0\|_1\sup_{1\le j,l\le p}\sup_{|\tau-\tau_0|<\frac{v}{n^{1-2\varphi}}}\left|\sum_{i=1}^n X_i^{(j)}X_i^{(l)}[1(Q_i<\tau_0)-1(Q_i<\tau)]\right|\|\widehat{\delta}(\widehat{\tau})-\delta_0\|_1\\
\le&C(\frac{v}{n^{1-2\varphi}})^{\varpi}n(\frac{s_0\sqrt{\log p}}{\sqrt{n}})^2+C(\frac{v}{n^{1-2\varphi}})^{\varpi}n\frac{s_0\sqrt{\log p}}{\sqrt{n}}\|\delta_0\|_1.
\end{aligned}\tag{A.9.13}$$

Thus, $\Upsilon(v)\rightsquigarrow 0$. Combine with Lemma 24,

$$\Delta_n(v)\rightsquigarrow\Delta(v).$$

□

Lemma 27. *Under Assumption 1,2 and 8,*

$$n^{1-2\varphi}(\widehat{\tau}-\tau_0)\xrightarrow{d}\omega T,$$

where $\omega=\frac{\delta_0^{*\prime}E\left[X_iX_i'U_i^2|Q_i=\tau\right]\delta_0^*}{(\delta_0^{*\prime}E\left[X_iX_i'|Q_i=\tau_0\right]\delta_0^*)^2}$ *and* $T=\text{argmax}_r\left[-\frac{|r|}{2}+W(r)\right]$.
$W(r)$ *is defined as a two-sided Brownian motion on the real line,*

$$W(r)=\begin{cases}W_1(r), & \text{if } r\geq 0,\\ W_2(r), & \text{if } r<0.\end{cases}$$

where $W_1(r)$ *and* $W_2(r)$ *are independent standard Brownian motions on* $[0,\infty)$.

Proof. By Theorem 2,

$$n^{1-2\varphi}(\widehat{\tau}-\tau_0)\leq C\frac{s_0\log p}{n^{2\varphi}}=O_p(1)$$

and by Lemma 26

$$\Delta_n(v)\rightsquigarrow\Delta(v).$$

Next, as $\lim_{v\to\infty}\frac{W(v)}{v}=0$, $\lim_{|v|\to\infty}\Delta(v)=-\infty$. Then the limit functional $\Delta(v)$ is continuous, so $Q(v)$ has a unique maximum. Therefore, all conditions of Theorem 2.7 of Kim and Pollard (1990) are satisfied, which implies

$$n^{1-2\varphi}(\widehat{\tau}-\tau_0)\xrightarrow{d}\text{argmax}_v Q(v).$$

Making the change-of-variables $v=\frac{\delta_0^{*\prime}E\left[X_iX_i'U_i^2|Q_i=\tau\right]\delta_0^*}{(\delta_0^{*\prime}E\left[X_iX_i'|Q_i=\tau_0\right]\delta_0^*)^2}r$, I can re-write the asymptotic distribution as

$$\text{argmax}_v Q(v)=\frac{\delta_0^{*\prime}E\left[X_iX_i'U_i^2|Q_i=\tau\right]\delta_0^*}{(\delta_0^{*\prime}E\left[X_iX_i'|Q_i=\tau_0\right]\delta_0^*)^2}\text{argmax}_r\left[-\frac{|r|}{2}+W(r)\right]$$

□

To test hypothesis $H_0:\tau=\tau_0$, a standard approach is to use the likelihood ratio statistic. Let $LR_n(\tau)=n\frac{S_n(\widehat{\alpha}(\tau),\tau)+\lambda\|\boldsymbol{D}(\tau)\widehat{\alpha}(\tau)\|_1-S_n(\widehat{\alpha}(\widehat{\tau}),\widehat{\tau})-\lambda\|\boldsymbol{D}(\widehat{\tau})\widehat{\alpha}(\widehat{\tau})\|_1}{S_n(\widehat{\alpha}(\widehat{\tau}),\widehat{\tau})+\lambda\|\boldsymbol{D}(\widehat{\tau})\widehat{\alpha}(\widehat{\tau})\|_1}$

Lemma 28. *Under Assumption 1,2 and 8,*

$$LR_n(\tau) \xrightarrow{d} \varrho^2 \Lambda,$$

where $\varrho^2 = \frac{\delta_0^{*\prime} E\left[X_i X_i' U_i^2 | Q_i = \tau\right]\delta_0^*}{\delta_0^{*\prime} E\left[X_i X_i' | Q_i = \tau_0\right]\delta_0^* \sigma^2}$ *and* $\Lambda = max_r\left[-\frac{|r|}{2} + W(r)\right]$.
The distribution function of Λ *is* $\mathbb{P}\{\Lambda < x\} = (1 - e^{-\frac{x^2}{2}})^2$.

Proof. We note that

$$S_n(\widehat{\alpha}(\widehat{\tau}), \widehat{\tau}) = \frac{1}{n}\sum_{i=1}^{n} \widehat{U}_i(\widehat{\tau})^2$$

and

$$\lambda \|\boldsymbol{D}(\widehat{\tau})\widehat{\alpha}(\widehat{\tau})\|_1 \xrightarrow{p} 0.$$

$$\begin{aligned}
&(S_n(\widehat{\alpha}(\widehat{\tau}), \widehat{\tau}) + \lambda \|\boldsymbol{D}(\widehat{\tau})\widehat{\alpha}(\widehat{\tau})\|_1) LR_n(\tau_0) - \Delta_n(\widehat{v}) \\
&= S_n(\widehat{\alpha}(\tau_0), \tau_0) + \lambda \|\boldsymbol{D}(\tau_0)\widehat{\alpha}(\tau_0)\|_1 - S_n(\widehat{\alpha}(\widehat{\tau}), \tau_0) - \lambda \|\boldsymbol{D}(\widehat{\tau})\widehat{\alpha}(\tau_0)\|_1 \qquad \text{(A.9.14)}\\
&= (\widehat{\alpha}(\tau_0) - \widehat{\alpha}(\widehat{\tau}))' \boldsymbol{X}(\tau_0)' \boldsymbol{X}(\tau_0)(\widehat{\alpha}(\tau_0) - \widehat{\alpha}(\widehat{\tau})) \xrightarrow{p} 0.
\end{aligned}$$

$$\text{(A.9.15)} \quad LR_n(\tau_0) = \frac{\Delta_n(n^{1-2\varphi}(\tau_0 - \widehat{\tau}))}{S_n(\widehat{\alpha}(\widehat{\tau}), \widehat{\tau}) + \lambda \|\boldsymbol{D}(\widehat{\tau})\widehat{\alpha}(\widehat{\tau})\|_1} = \frac{Sup_v \Delta_n(v)}{S_n(\widehat{\alpha}(\widehat{\tau}), \widehat{\tau}) + \lambda \|\boldsymbol{D}(\widehat{\tau})\widehat{\alpha}(\widehat{\tau})\|_1} \xrightarrow{d} \frac{Sup_v \Delta(v)}{\sigma^2}$$

by continuous mapping theorem. This limiting distribution equals

$$\begin{aligned}
&\frac{1}{\sigma^2} Sup_v \left[-v\delta_0^{*\prime} E\left[X_i X_i' | Q_i = \tau_0\right]\delta_0^* + 2\sqrt{|v|\delta_0^{*\prime} E\left[X_i X_i' U_i^2 | Q_i = \tau\right]\delta_0^*} W(v)\right] \qquad \text{(A.9.16)}\\
&= \frac{\delta_0^{*\prime} E\left[X_i X_i' U_i^2 | Q_i = \tau\right]\delta_0^*}{\delta_0^{*\prime} E\left[X_i X_i' | Q_i = \tau_0\right]\delta_0^* \sigma^2} \sup_r \left[-\frac{|r|}{2} + W(r)\right] = \varrho^2 \Lambda
\end{aligned}$$

To find the distribution function of Λ, note that

$$\sup_r \left[-\frac{|r|}{2} + W(r)\right] = 2\max\left[\sup_{r>0}[-\frac{|r|}{2} + W(r)], \sup_{r<0}[-\frac{|r|}{2} + W(r)]\right] = 2\max[\Lambda_+, \Lambda_-].$$

which becomes the two-sided Brownian motion, as in Hansen (2000). $[\Lambda_+, \Lambda_-]$ are iid exponential random variables with distribution $\mathbb{P}\{\Lambda_+ < x\} = 1 - e^{-x}$. see Bhattacharya and Brockwell 1976.

Thus

$$\mathbb{P}\{\Lambda < x\} = \mathbb{P}\{2\max[\Lambda_+, \Lambda_-] < x\} = \mathbb{P}\{2\Lambda_+ < x\}\mathbb{P}\{2\Lambda_- < x\} = (1 - e^{-\frac{x^2}{2}})^2.$$

$\square$

The likelihood ratio test corresponds to a modified version of the LR Test used in Hansen (2000). The asymptotic distribution of Lemma 28 depends on the nuisance parameter ϱ^2, which can be constructed by following Section 3.4 in Hansen (2000). We can then obtain a confidence interval for τ.

APPENDIX

B

CHAPTER 2

B.1 Tables

Augmented Dickey-Fuller test Results		
Variable		ADF
Unit Root		
ln_US_Ukraine_diff		-4.446
ln_US_Argentina_diff		-3.860
ln_Ukraine_Argentina_diff		-4.877
Alternative hypothesis: stationary	Lag order = 6	
significance level	Critical value	
1%	-3.96	
5%	-3.41	
10%	-3.12	

*The critical values are interpolated from Table 4.2 of Banerjee et al. (1993).

Table B.1: Augmented Dickey-Fuller Test Results of Price Differentials

Augmented Dickey-Fuller Test		
Variable (1st diff)		Dickey-Fuller
ln_US_Ukraine_diff		-5.403
ln_US_Argentina_diff		-8.379
ln_Ukraine_Argentina_diff		-6.283
ln_USD_UAH_Exchange		-5.207
ln_USD_Peso_Exchange		-5.222
ln_Peso_UAH_Exchange		-4.650
ln_Baltic_Freight		-8.028
US Unemployment Rate		-7.146
Ukraine Unemployment Rate		-6.261
Argentina Unemployment Rate		-5.148
ln_US Industrial Production Index		-5.794
ln_Ukraine Industrial Production Index		-6.467
ln_Argentina Industrial Production Index		-7.630
ln_US Consumer price index		-6.301
ln_Ukraine Consumer price index		-4.290
Argentina Monthly Inflation		-6.374
US Interest Rate		-3.055
ln_US Monthly Gas Price		-7.502
ln_US Corn Stock		-11.561
Alternative hypothesis: stationary	Lag order = 6	
significance level	Critical value	
1%	-3.96	
5%	-3.41	
10%	-3.12	

*The critical values are interpolated from Table 4.2 of Banerjee et al. (1993).

Table B.2: Augmented Dickey-Fuller Test Results of First Difference of Time Series

time delay for the threshold variable	1	2	3	4	5	6
US/ Ukraine						
BIC	-3.307	-4.779	-3.582	-5.087	-3.635	-4.227
Threshold estimate	0.11	0.15	0.07	0.09	0.05	0.11
Threshold estimate(quantile)	0.64	0.77	0.42	0.55	0.27	0.67
US/Argentina						
BIC	-4.932	-5.106	-5.111	-6.515	-6.453	-6.474
Threshold estimate	0.06	0.05	0.07	0.04	0.07	0.07
Threshold estimate(quantile)	0.58	0.52	0.68	0.46	0.70	0.70
Ukraine/Argentina						
BIC	-5.036	-4.997	-5.177	-4.774	-4.269	-4.597
Threshold estimate	0.11	0.12	0.13	0.12	0.13	0.11
Threshold estimate(quantile)	0.70	0.73	0.77	0.74	0.77	0.69

Table B.3: LASSO Estimation with BIC

Model	US/Ukraine	US/Argentina	Ukraine/Argentina
Baltic Dry Index	-		-
US Unemployment Rate	-		
lag 1 Baltic Dry Index			-
lag 5 Argentina Inflation			-
lag 2 Argentina INDPRO			+
trade lag 4 Argentina Inflation			-

Table B.4: Signs of Market Control Variables Estimates

time delay for the threshold variable	1	2	3	4	5	6
Entire Period						
BIC	-3.307	-4.779	-3.582	-5.087	-3.635	-4.227
Threshold estimate	0.11	0.15	0.07	0.09	0.05	0.11
Threshold estimate(quantile)	0.64	0.77	0.42	0.55	0.27	0.67
Pre-October 2008						
BIC	-5.805	-5.291	-5.493	-5.956	-5.610	-6.009
Threshold estimate	0.11	0.09	0.12	0.13	0.21	0.24
Threshold estimate(quantile)	0.58	0.44	0.62	0.66	0.76	0.79
Post-October 2008						
BIC	-5.122	-3.907	-4.215	-3.817	-3.753	-3.919
Threshold estimate	0.11	0.08	0.08	0.06	0.06	0.06
Threshold estimate(quantile)	0.68	0.53	0.49	0.36	0.36	0.36
Pre-February 2014						
BIC	-3.625	-3.096	-3.514	-3.327	-3.677	-4.522
Threshold estimate	0.10	0.08	0.12	0.09	0.07	0.14
Threshold estimate(quantile)	0.58	0.48	0.68	0.49	0.39	0.73
Post-February 2014						
BIC	-5.529	-5.545	-5.713	-5.739	-5.690	-5.077
Threshold estimate	0.09	0.07	0.06	0.06	0.06	0.05
Threshold estimate(quantile)	0.56	0.44	0.39	0.37	0.45	0.32

Table B.5: LASSO Estimation with BIC: US/ Ukraine

time delay for the threshold variable	1	2	3	4	5	6
Entire Period						
BIC	-4.932	-5.106	-5.111	-6.515	-6.453	-6.474
Threshold estimate	0.06	0.05	0.07	0.04	0.07	0.07
Threshold estimate(quantile)	0.58	0.52	0.68	0.46	0.70	0.70
Pre-October 2008						
BIC	-6.830	-7.180	-7.573	-7.753	-6.869	-7.202
Threshold estimate	0.06	0.06	0.06	0.04	0.05	0.05
Threshold estimate(quantile)	0.56	0.52	0.54	0.38	0.45	0.45
Post-October 2008						
BIC	-4.124	-4.691	-4.904	-4.745	-4.806	-5.323
Threshold estimate	0.05	0.05	0.06	0.06	0.07	0.05
Threshold estimate(quantile)	0.52	0.55	0.59	0.59	0.62	0.55

Table B.6: LASSO Estimation with BIC: US/ Argentina

time delay for the threshold variable	1	2	3	4	5	6
Entire Period						
BIC	-5.036	-4.997	-5.177	-4.774	-4.269	-4.597
Threshold estimate	0.11	0.12	0.13	0.12	0.13	0.11
Threshold estimate(quantile)	0.70	0.73	0.77	0.74	0.77	0.69
Pre-February 2014						
BIC	-3.268	-3.941	-4.038	-4.195	-3.953	-3.548
Threshold estimate	0.10	0.09	0.13	0.11	0.08	0.08
Threshold estimate(quantile)	0.62	0.56	0.67	0.61	0.51	0.46
Post-February 2014						
BIC	-5.646	-5.514	-5.539	-5.828	-4.912	-4.946
Threshold estimate	0.09	0.05	0.06	0.07	0.09	0.06
Threshold estimate(quantile)	0.70	0.48	0.58	0.63	0.69	0.58

Table B.7: LASSO Estimation with BIC: Ukraine/Argentina

Parameter	Entire Period	Pre-October 2008	Post-October 2008	Pre-February 2014	Post-February 2014
(Intercept) γ_0	-0.018**	-0.026***	0.025	-0.011 **	-0.032***
	(0.008)	(0.000)	(0.030)	(0.005)	(0.001)
degree of "error correction" γ_1	-0.090***		-0.080 ***	-0.079***	-0.390***
	(0.023)		(0.016)	(0.027)	(0.039)
exchange rate effect $\gamma_{2,0}$	-0.023	-1.321**	0.163		0.069***
	(0.094)	(0.553)	(0.126)		(0.004)
(Intercept) δ_0					
degree of "error correction" δ_1	-0.070 *	-0.237***	0.087	0.000	0.200***
	(0.041)	(0.004)	(0.204)	(0.086)	(0.059)
exchange rate effect $\delta_{2,0}$	0.300	-0.565 ***			0.572 ***
	(0.223)	(0.168)			(0.054)
other variables omitted	…	…	…	…	…
	…	…	…	…	…
threshold estimate	0.089	0.237	0.111	0.143	0.056
threshold quantile	0.55	0.79	0.68	0.73	0.37
optimal threshold time delay	4	6	1	6	4
Observations	233	73	160	137	96

Note: *p<0.1; **p<0.05; ***p<0.01

Table B.8: Estimates of Threshold Model Using Debiased LASSO: US/ Ukraine

Parameter	Entire Period	Pre-October 2008	Post-October 2008
(Intercept) γ_0	0.009	-0.005***	-0.011*
	(0.008)	(0.000)	(0.006)
degree of “error correction” γ_1		-0.097***	-0.153**
		(0.000)	(0.061)
exchange rate effect $\gamma_{2,0}$			0.238***
			(0.005)
(Intercept) δ_0		-0.022***	-0.023*
		(0.001)	(0.013)
degree of “error correction” δ_1	-0.160***	-0.117***	0.016
	(0.023)	(0.010)	(0.080)
exchange rate effect $\delta_{2,0}$		1.465***	0.261*
		(0.233)	(0.141)
	…	…	…
	…	…	…
threshold estimate	0.043	0.040	0.051
threshold quantile	0.46	0.38	0.55
optimal threshold time delay	4	4	6
Observations	228	58	170

Note: *p<0.1; **p<0.05; ***p<0.01

Table B.9: Estimates of Threshold Model Using Debiased LASSO: US/Argentina

Parameter	Entire Period	Pre-February 2014	Post-February 2014
(Intercept) γ_0	0.087	0.105**	-0.043***
	(0.093)	(0.054)	(0.005)
degree of "error correction" γ_1	-0.139 ***	-0.569***	-0.555***
	(0.020)	(0.054)	(0.089)
exchange rate effect $\gamma_{2,0}$	0.473	1.169 ***	0.228***
	(0.410)	(0.060)	(0.027)
(Intercept) δ_0		-0.160	
		(0.135)	
degree of "error correction" δ_1		0.943***	-0.004
		(0.139)	(0.005)
exchange rate effect $\delta_{2,0}$	9.051	-3.617*	0.119***
	(8.442)	(1.880)	(0.004)
	…	…	…
	…	…	…
threshold estimate	0.111	0.106	0.071
threshold quantile	0.77	0.61	0.63
optimal threshold time delay	3	4	4
Observations	218	122	96

Note: *p<0.1; **p<0.05; ***p<0.01

Table B.10: Estimates of Threshold Model Using Debiased LASSO: Ukraine/Argentina